First experimental search for inclusive Higgs boson production using the $H \to b\bar{b}$ decay with CMS

Caterina Vernieri (FNAL & LPC Distinguished Researcher)
on behalf of the CMS experiment

Fermilab - Friday June 30, 2017 - Wine and Cheese Seminar

The Higgs boson

- A great advance in our understanding of fundamental particles and their interactions
    - a **new state** of matter-energy
    - a **potential window to Beyond Standard Model (SM)**

H^0 $\qquad J = 0$

Mass $m = 125.09 \pm 0.24$ GeV
Full width $\Gamma < 1.7$ GeV, CL = 95%

H^0 Signal Strengths in Different Channels

See Listings for the latest unpublished results.

Combined Final States $= 1.10 \pm 0.11$
$W W^* = 1.08^{+0.18}_{-0.16}$
$Z Z^* = 1.29^{+0.26}_{-0.23}$
$\gamma\gamma = 1.16 \pm 0.18$
$b\bar{b} = 0.82 \pm 0.30$ (S = 1.1)
$\mu^+\mu^- < 7.0$, CL = 95%
$\tau^+\tau^- = 1.12 \pm 0.23$
$Z\gamma < 9.5$, CL = 95%
$t\bar{t}H^0$ Production $= 2.3^{+0.7}_{-0.6}$

Outline

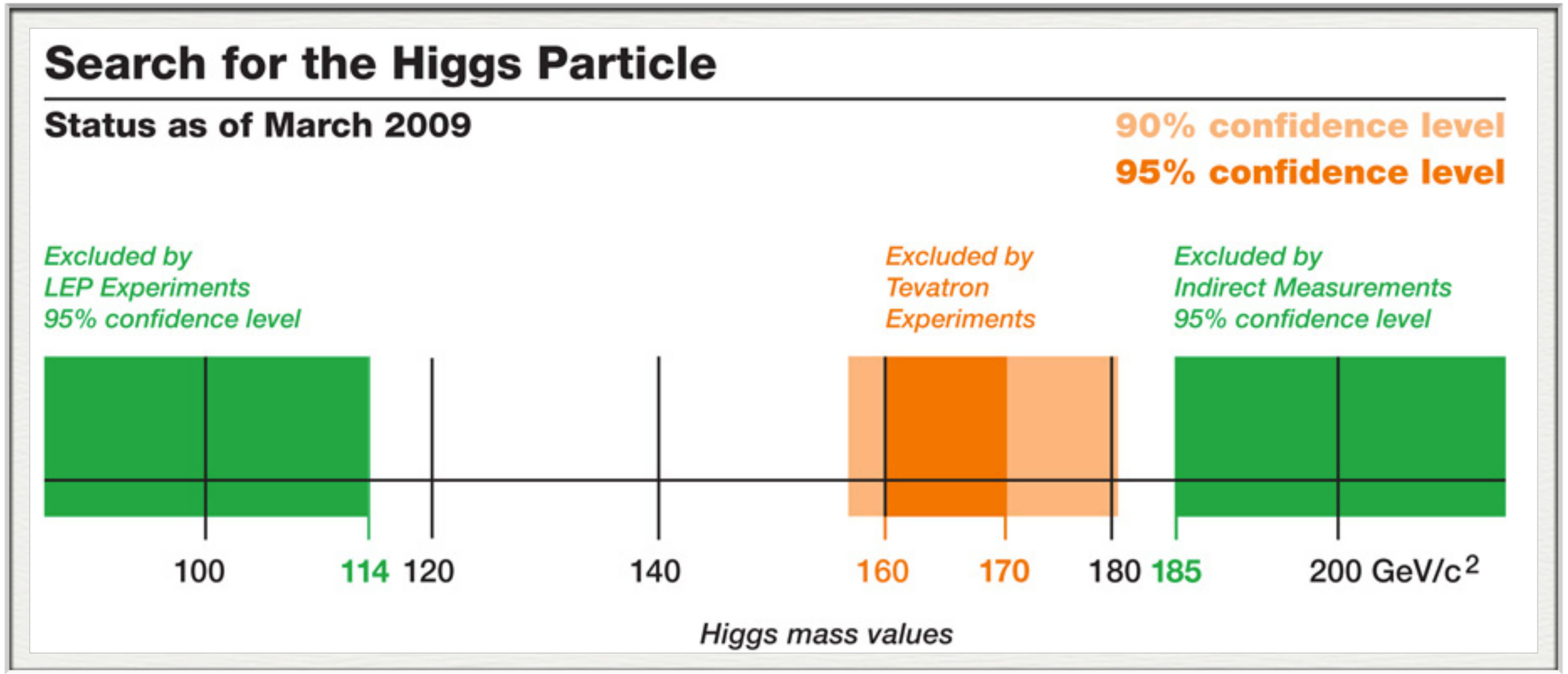
- **Higgs boson discovery at LHC**
- $H \rightarrow b\bar{b}$ state of the art
- Tools for identifying $H \rightarrow b\bar{b}$ at high p_T
    - b-quark identification in CMS
- Inclusive search for boosted $H \rightarrow b\bar{b}$
- Future perspectives

Standard Model Higgs Hunting: Basics

LEP+Tevatron legacy:

low-mass range [114,158] GeV

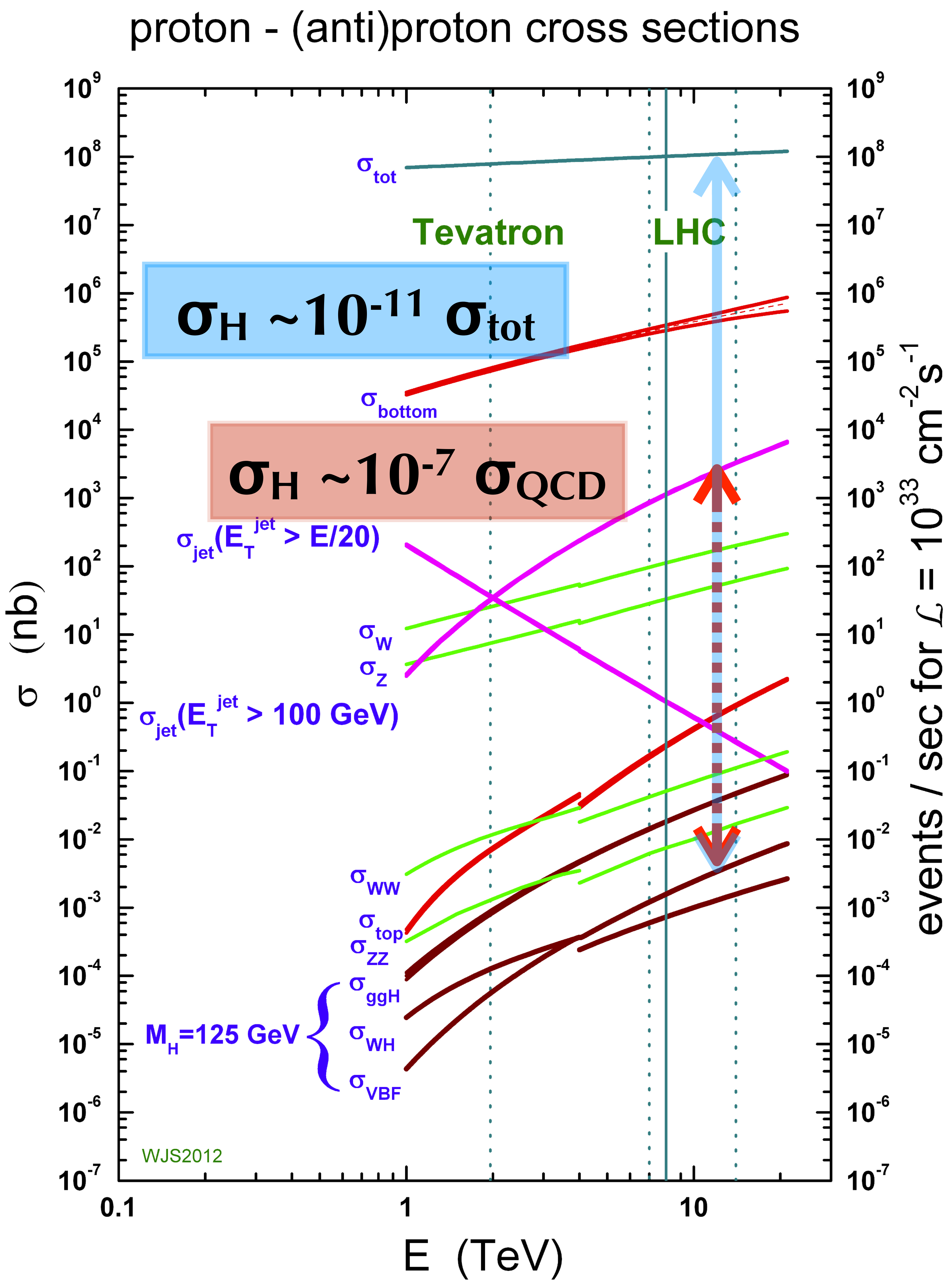
2009



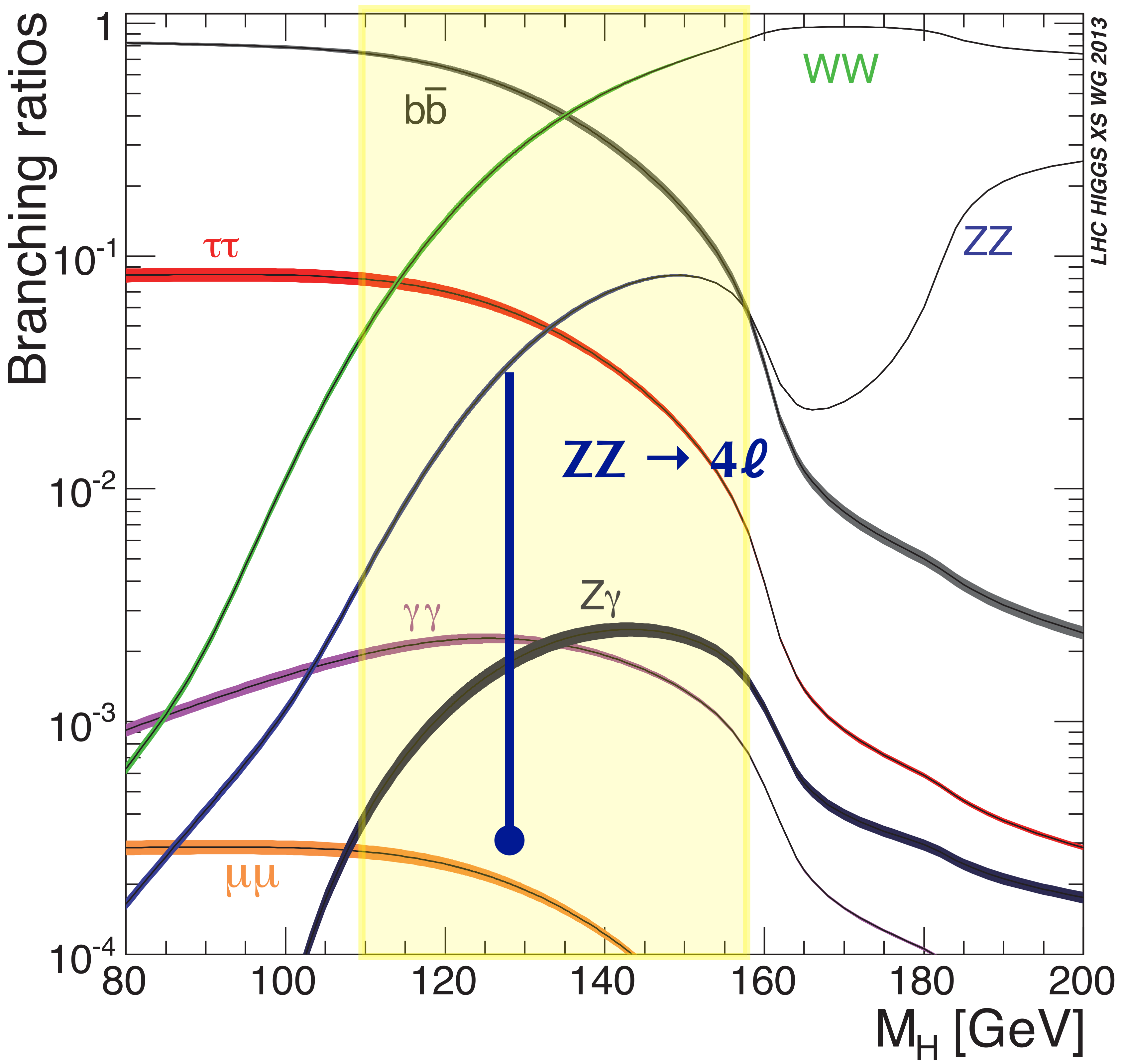
The natural width is less than 100 MeV

observed peak dominated by instrumental mass resolution

Standard Model Higgs Hunting: Strategy

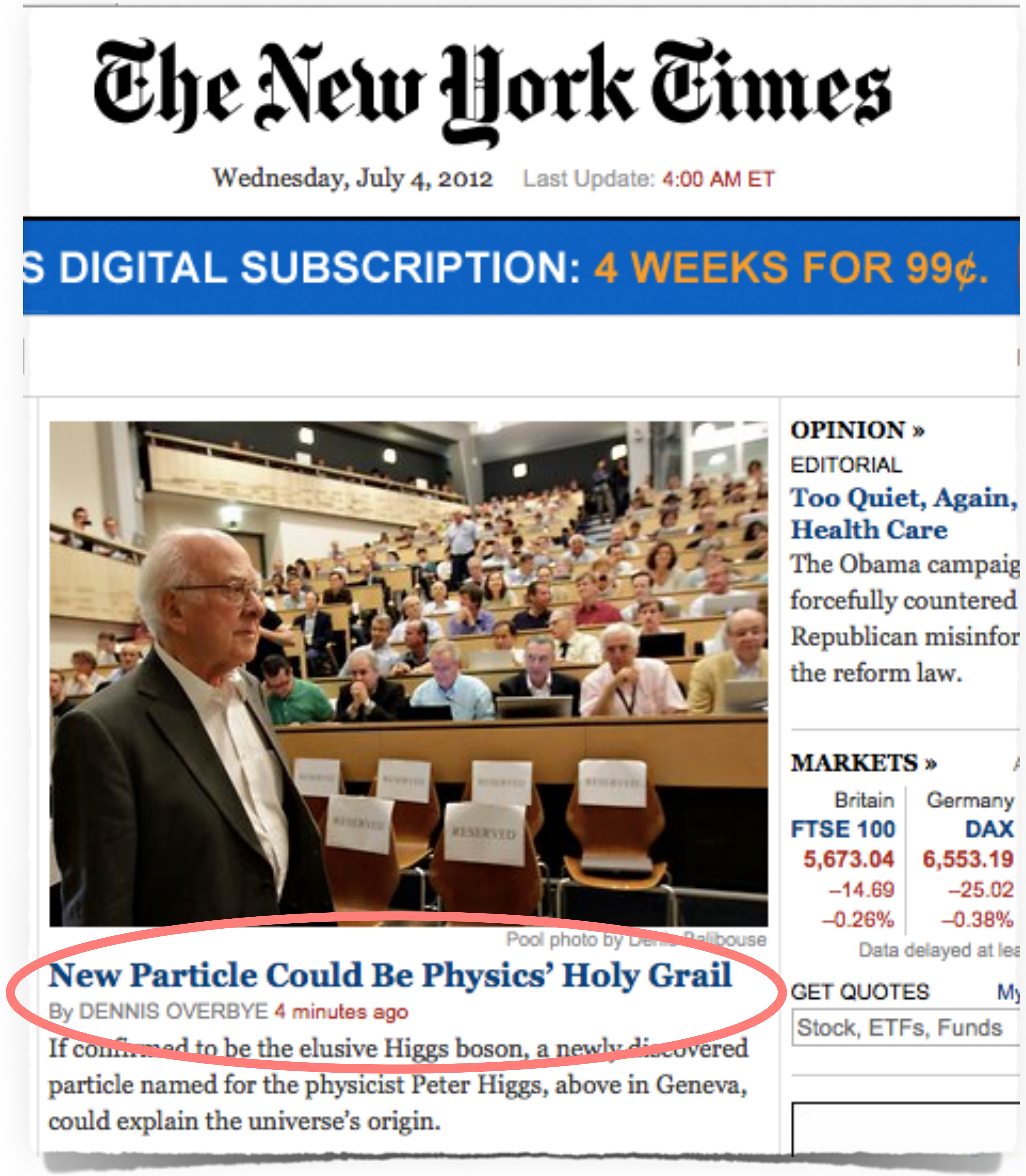


H needs { high energy
optimal invariant mass resolution
luminosity



lowest BR decay modes ↔
Excellent mass resolution

July 4th 2012, Higgs Boson discovery


The New York Times

Wednesday, July 4, 2012 Last Update: 4:00 AM ET

S DIGITAL SUBSCRIPTION: 4 WEEKS FOR 99¢.

Pool photo by Denis Balibouse

New Particle Could Be Physics' Holy Grail

By DENNIS OVERBYE 4 minutes ago

If confirmed to be the elusive Higgs boson, a newly discovered particle named for the physicist Peter Higgs, above in Geneva, could explain the universe's origin.

OPINION »

EDITORIAL

Too Quiet, Again, Health Care

The Obama campaig forcefully countered Republican misinfor the reform law.

MARKETS »

Britain	Germany
FTSE 100	DAX
5,673.04	6,553.19
−14.69	−25.02
−0.26%	−0.38%

Data delayed at lea

GET QUOTES My

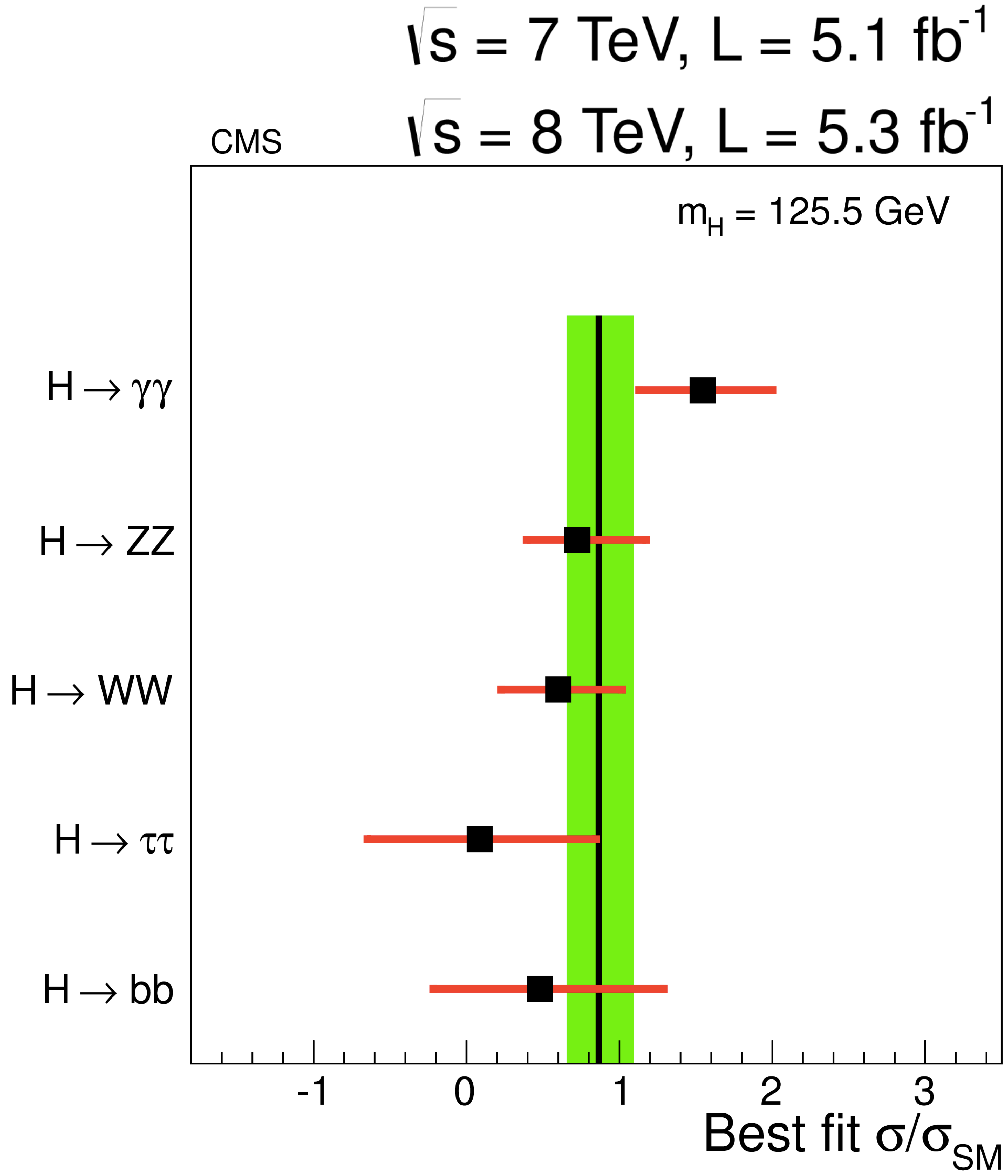
Stock, ETFs, Funds

Phys. Lett. B 716 (2012) 30

July 4th 2012, Higgs Boson discovery

CMS and ATLAS reported independently the **first observation** of the Higgs boson

- **5.0σ combining γγ and ZZ alone**
    - **best mass resolution**
    - thanks to the huge amount of LHC data we could exploit the lowest BR decay modes



Is it a SM Higgs boson?

- Mass
- Spin-parity
- Width
- **The couplings to fermions and bosons**
- Study the self-coupling
- Any non-SM property?

	Sensitivity (Run I+II)
$H \to b\bar{b}$	**$\sim 2\sigma$***
$H \to WW$	$> 5\sigma$
$H \to \tau\tau$	$\sim 5\sigma$
$H \to ZZ$	$\gg 5\sigma$
$H \to \gamma\gamma$	$\gg 5\sigma$

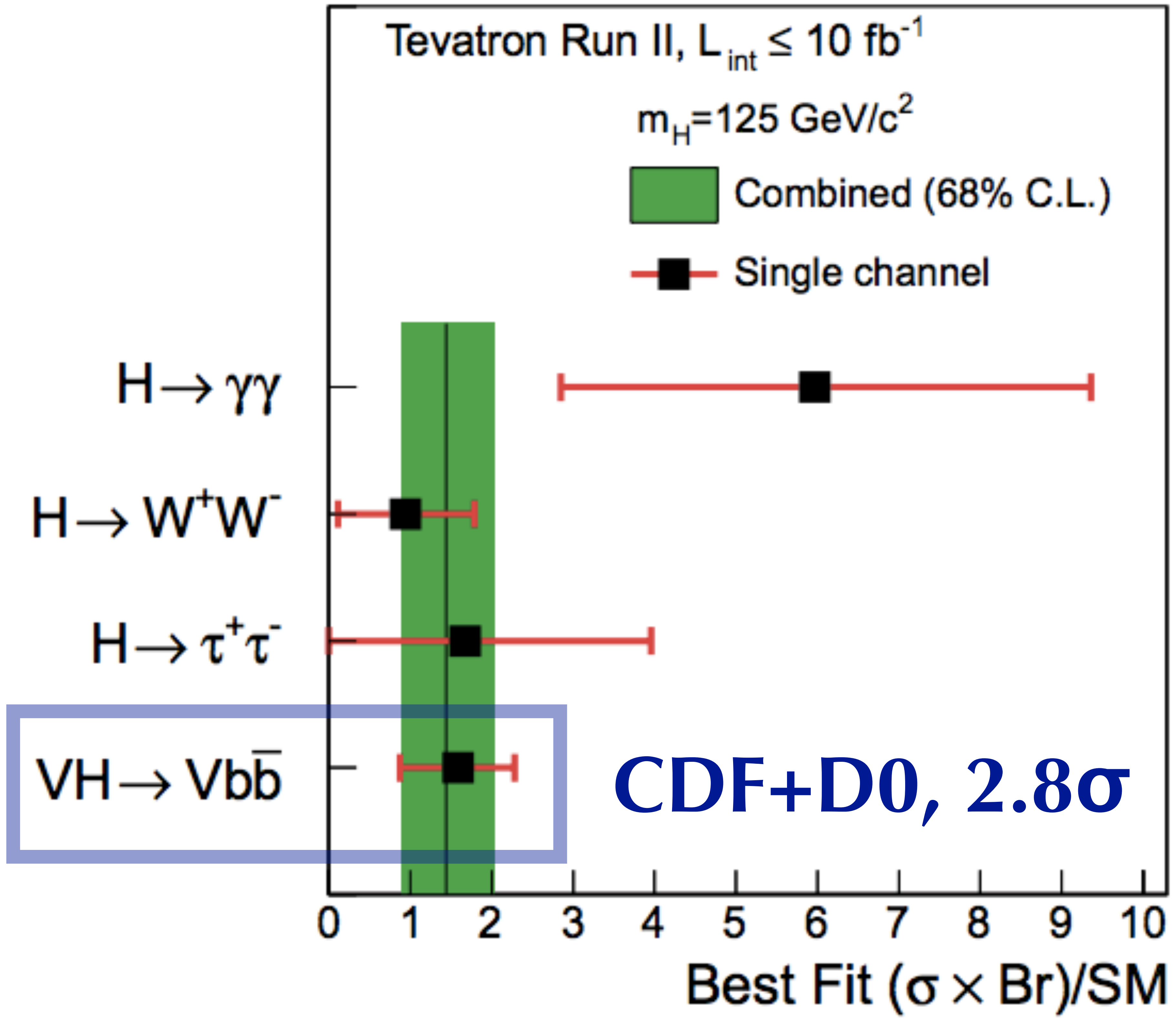
* Run I only

$b\bar{b}$, largest BR for SM H (~58%) but not yet observed

Higgs at Tevatron

CDF and D0 combined results reported a broad excess in the mass range $115 < m_H < 140$ GeV

- 3.0 **σ** at m_H=125 GeV
- **mainly from the H → $b\bar{b}$**



CDF+D0, 2.8σ

arXiv:1303.6346

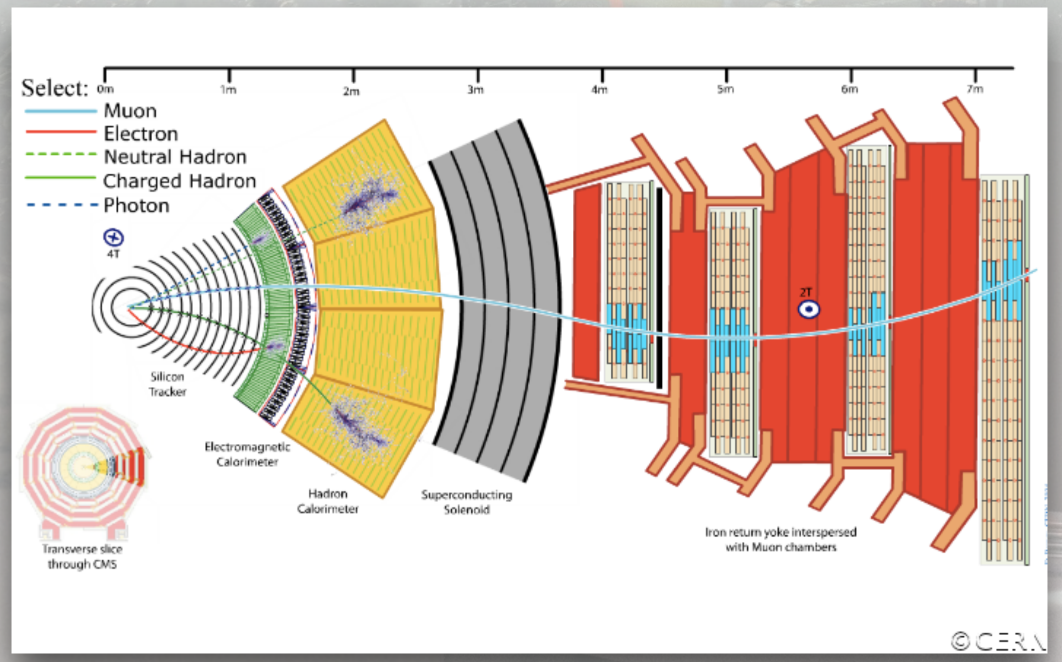
Outline

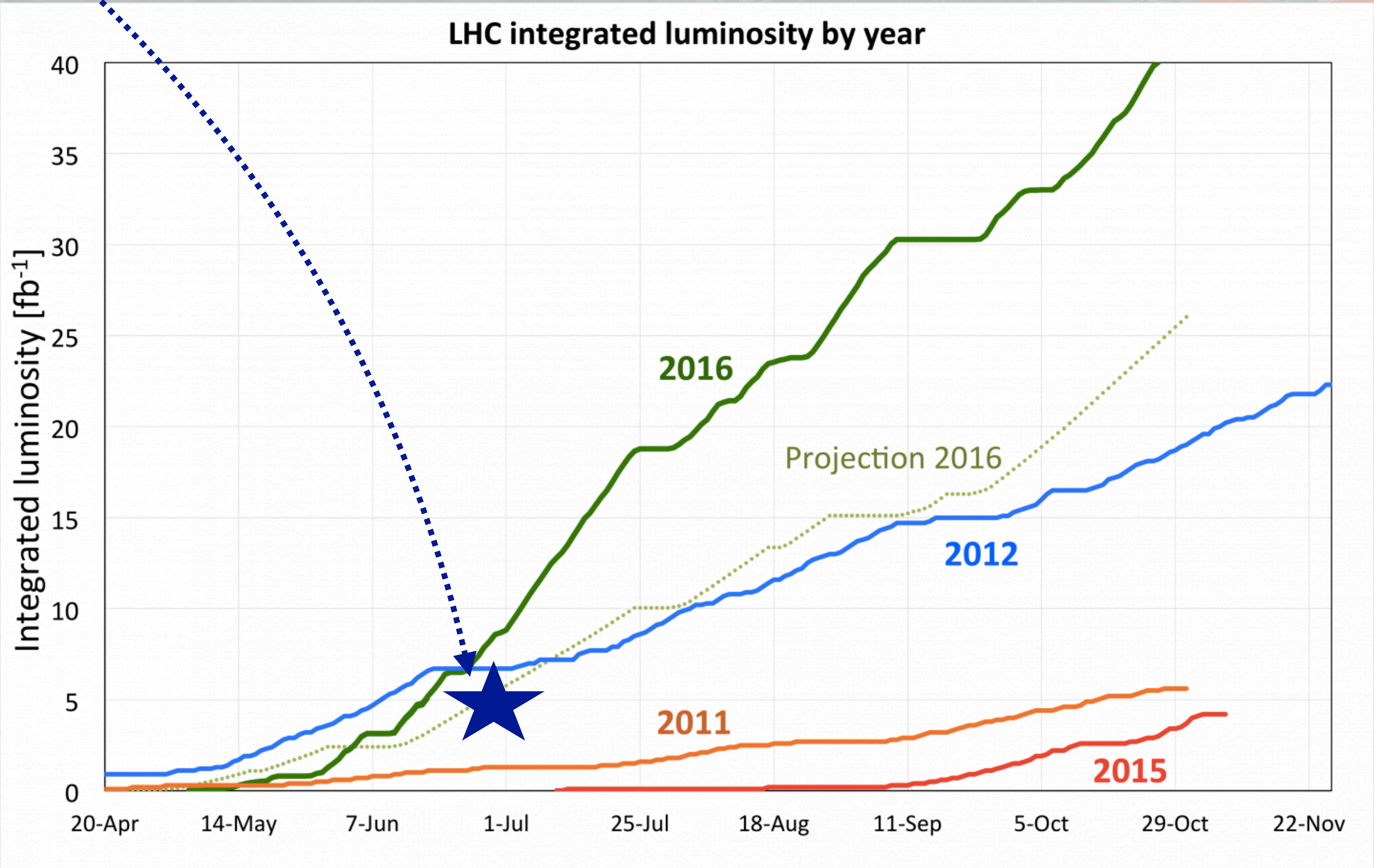
- Higgs boson discovery at LHC
- **$H \to b\bar{b}$ state of the art**
- Tools for identifying $H \to b\bar{b}$ at high p_T
  - b-quark identification in CMS
- Inclusive search for boosted $H \to b\bar{b}$
- Future perspectives

The Large Hadron Collider

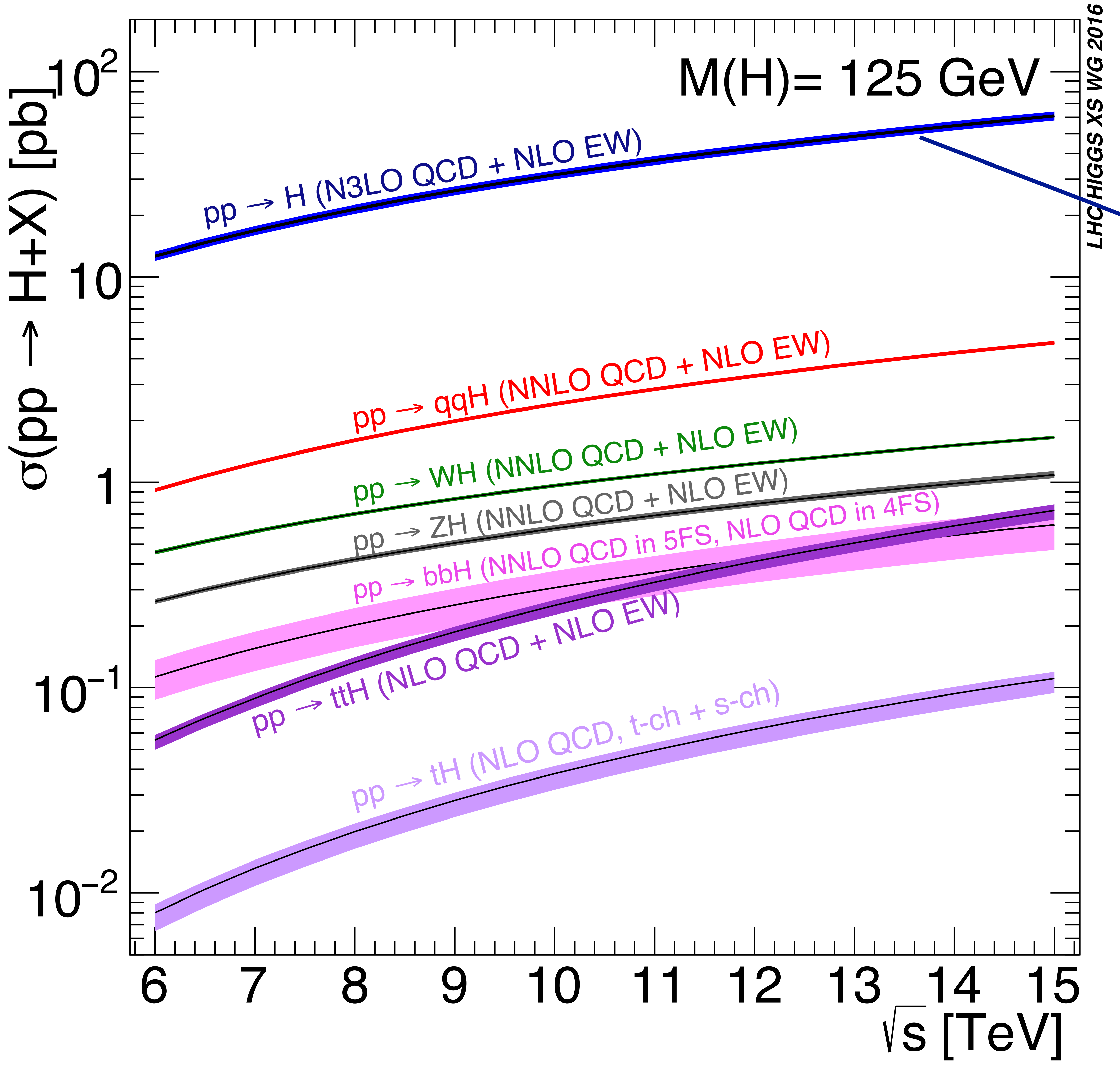
2010-2012, ~25fb^{-1} delivered in Run I at **7** and **8 TeV**
2015-2016, ~40 fb^{-1} delivered in Run II at **13 TeV**
2017, Run just started. So far CMS recorded >5 fb^{-1}

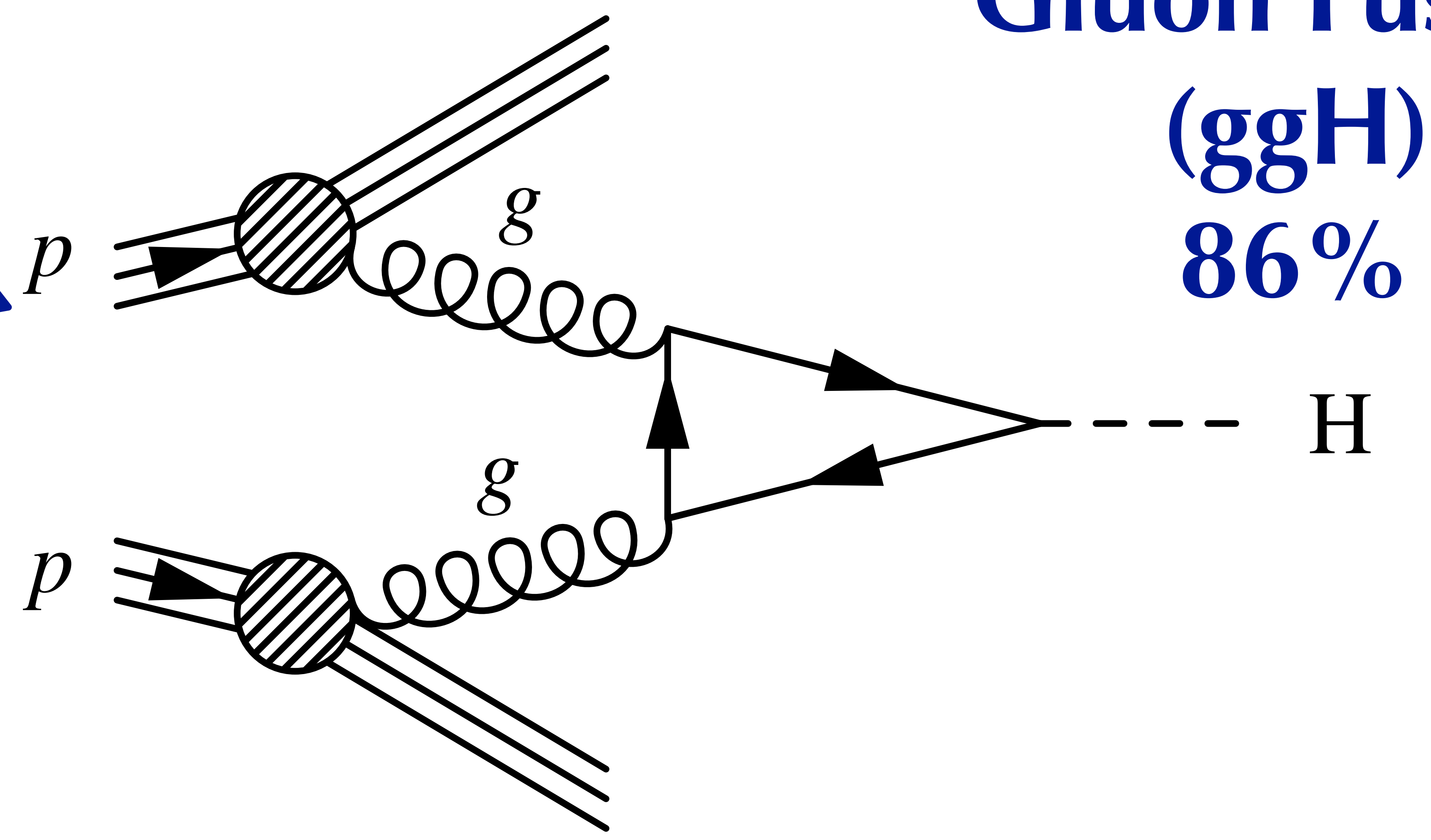
CMS





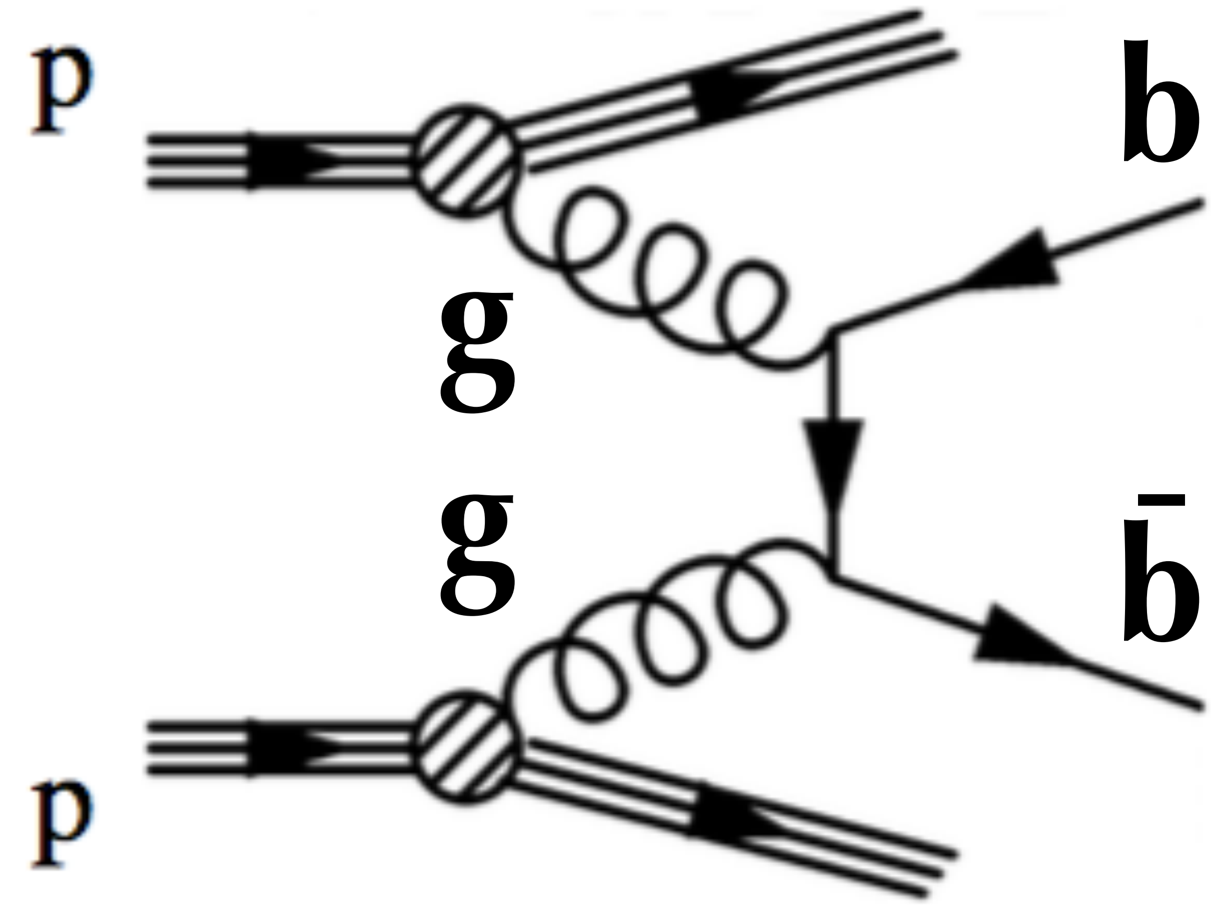
$H(b\bar{b})$ at LHC

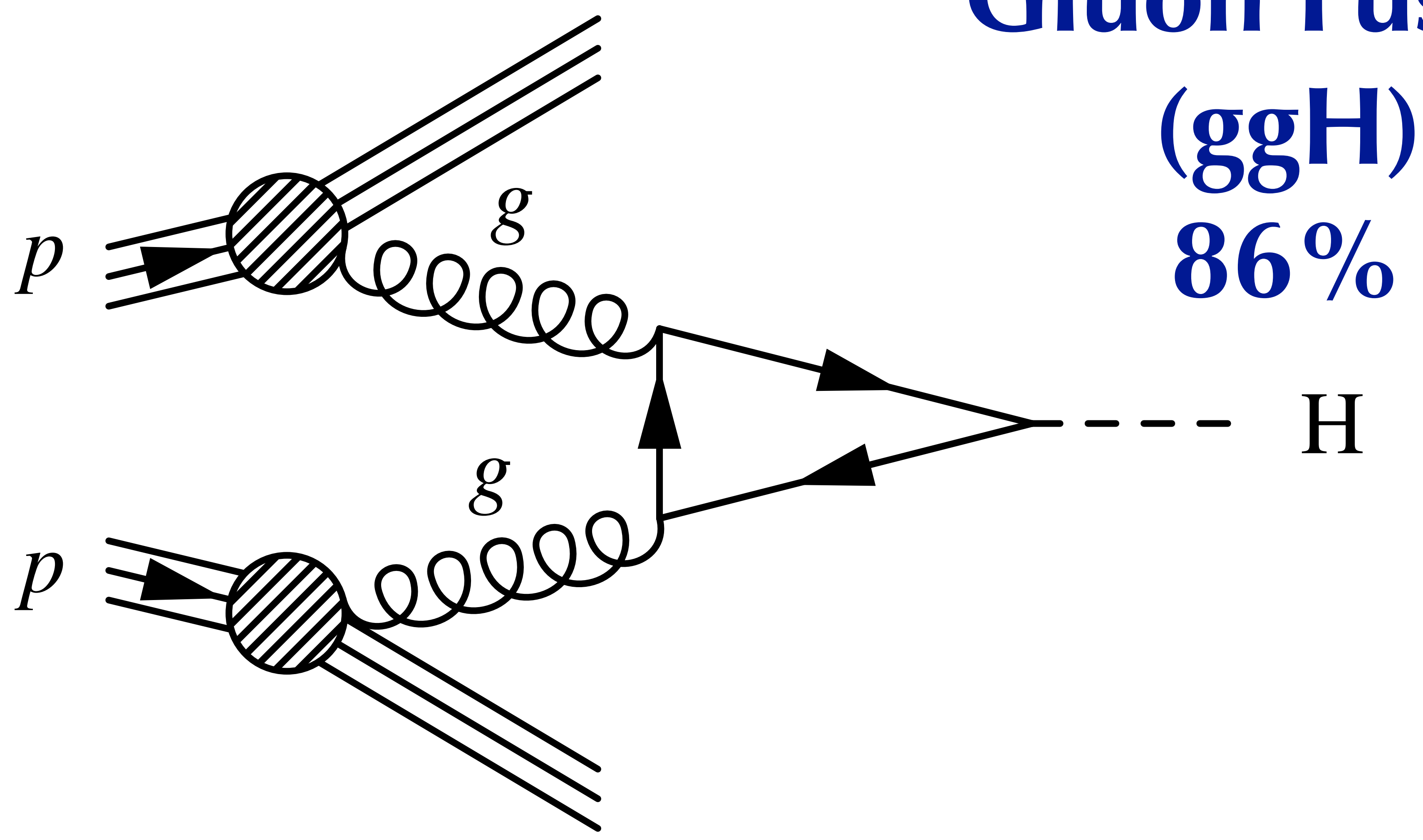




Gluon Fusion (ggH) 86%

$H(b\bar{b})$ at LHC



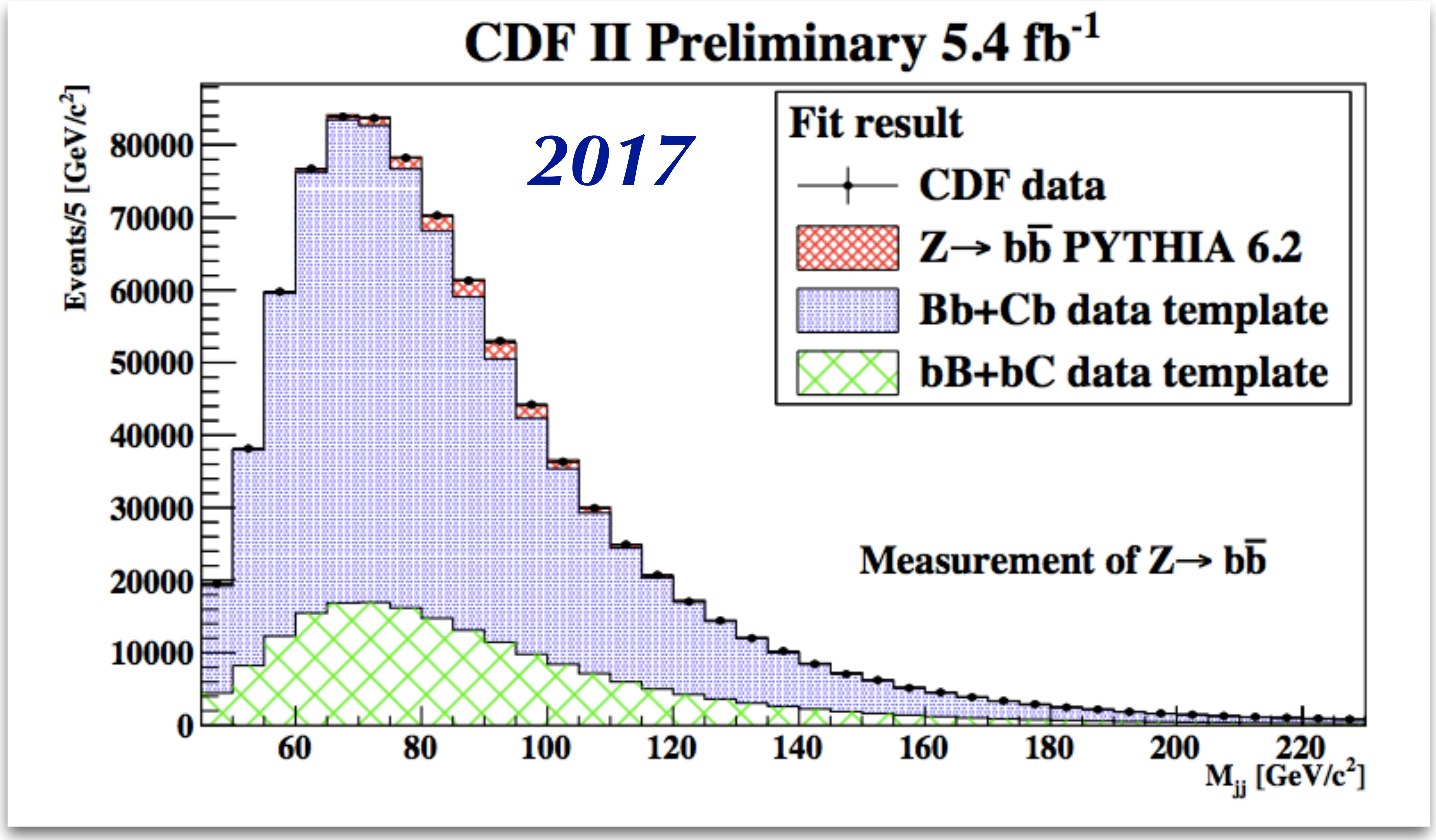


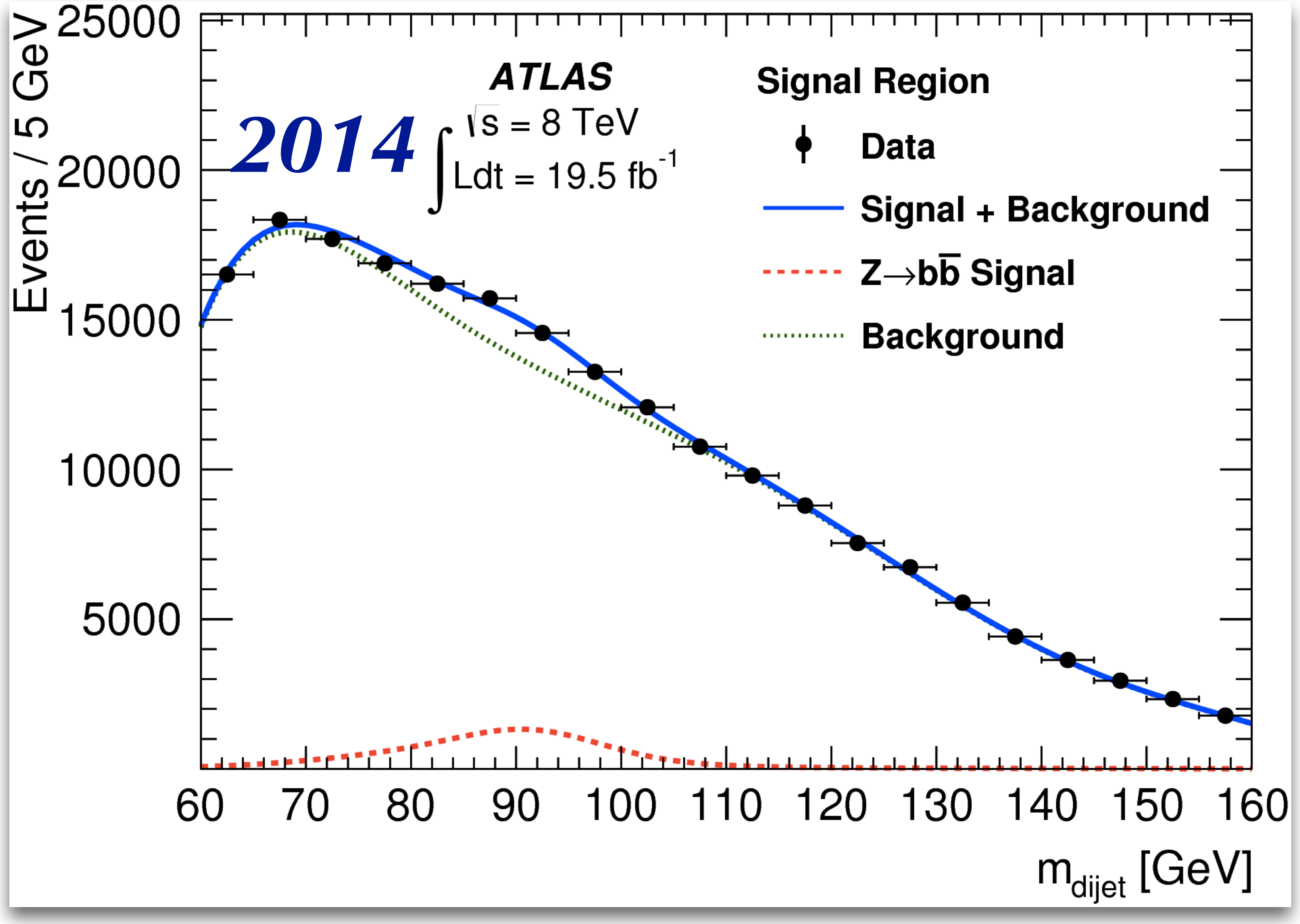
Gluon Fusion (ggH) 86%

- Two-jets final state
- Overwhelming background from QCD production of b quarks
  - 10^7 larger

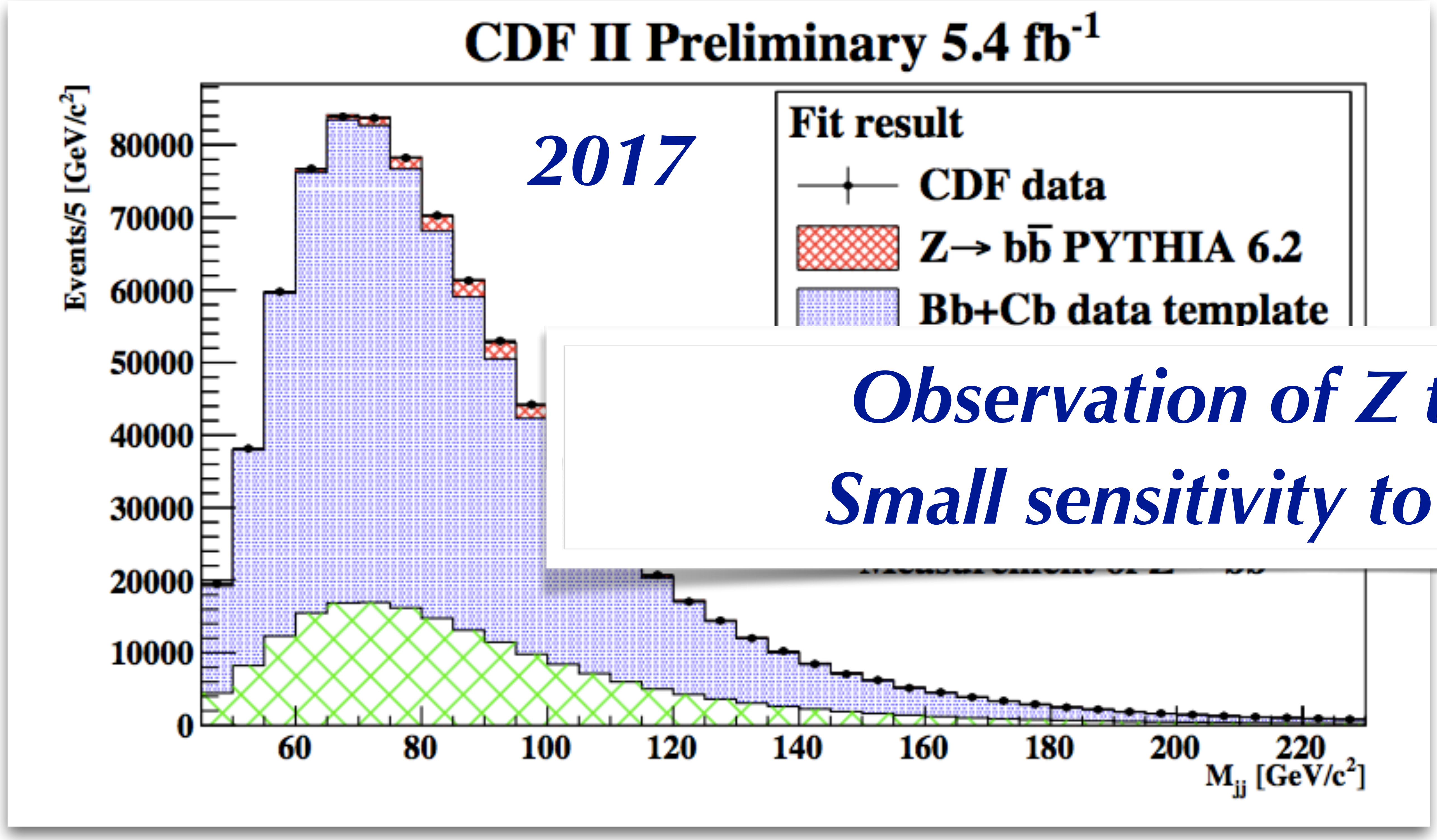
$H(b\bar{b})$ at LHC

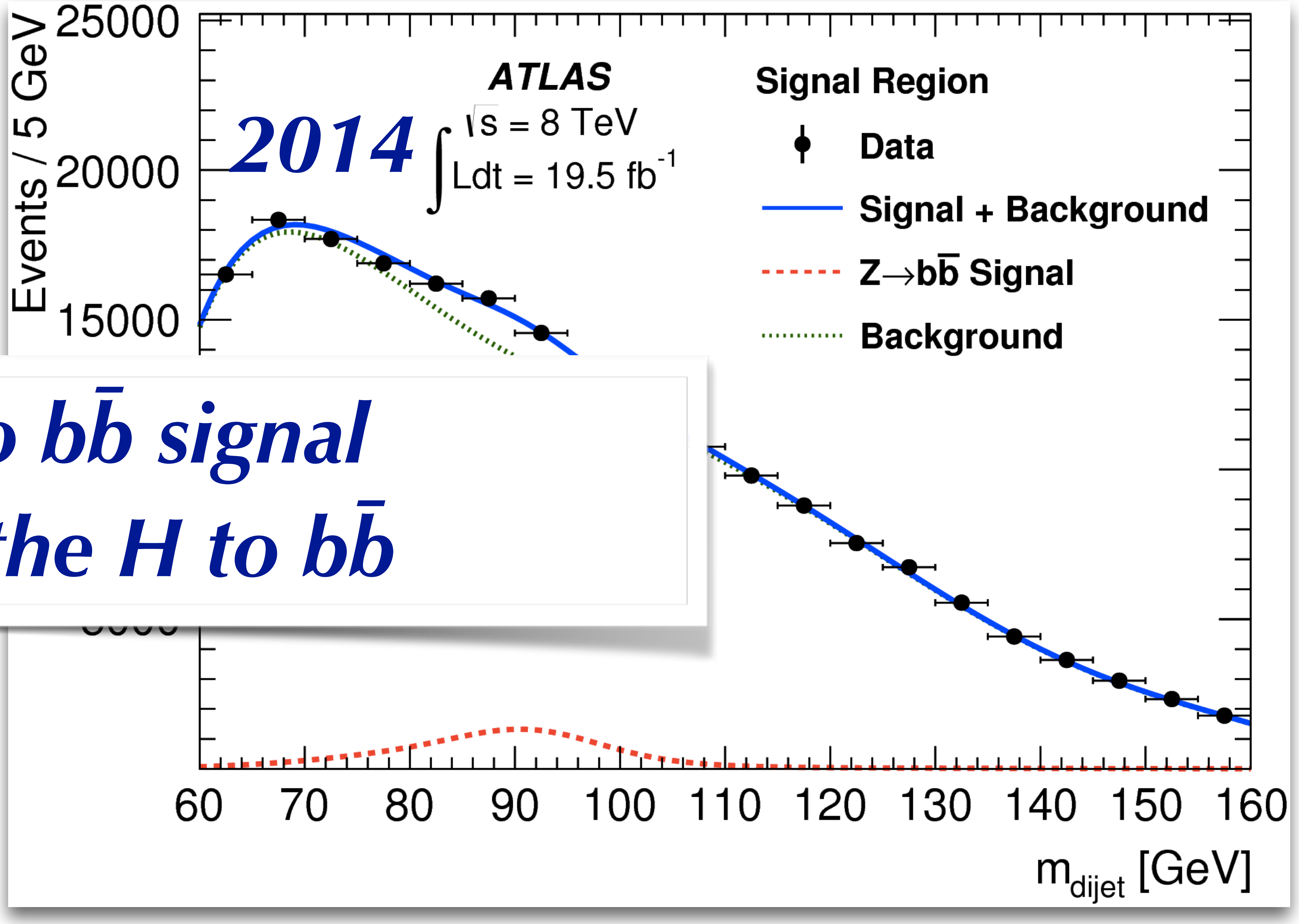
CDF-11228
Phys. Lett. B 738 (2014)





H($b\bar{b}$) at LHC

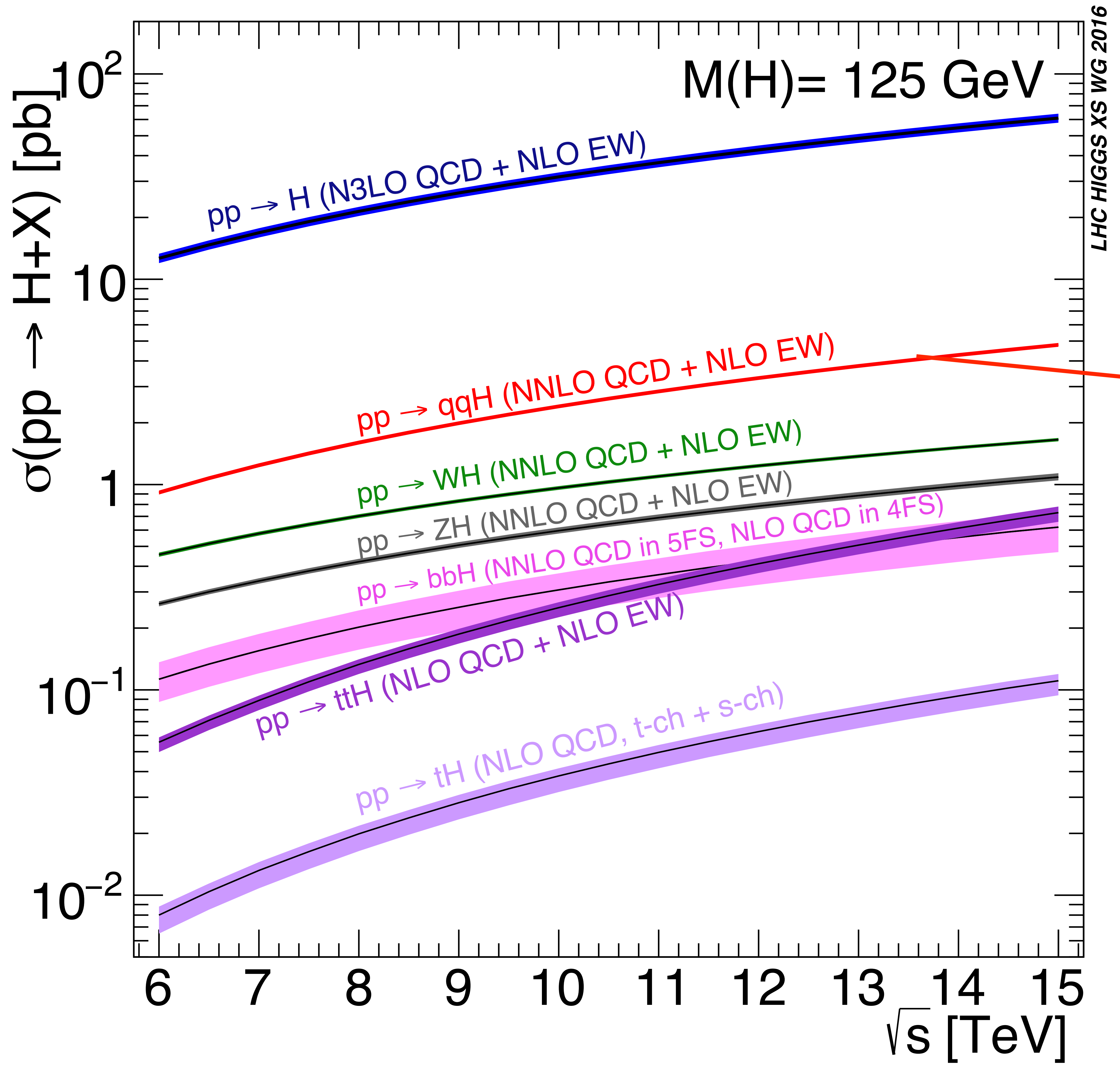




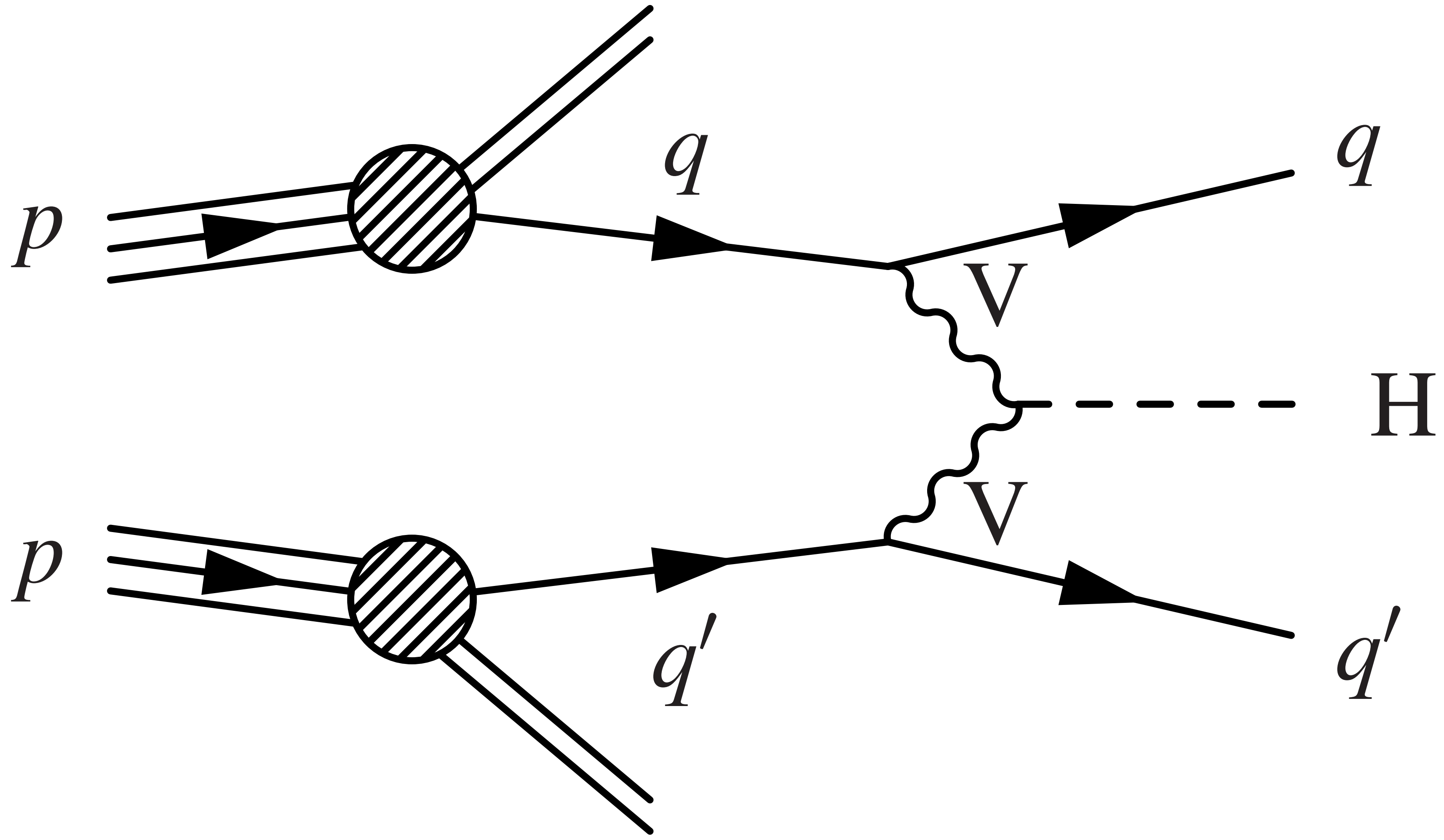
Observation of Z to $b\bar{b}$ signal
Small sensitivity to the H to $b\bar{b}$

Search for gg → H → $b\bar{b}$ historically **deemed impossible**

H($b\bar{b}$) at LHC

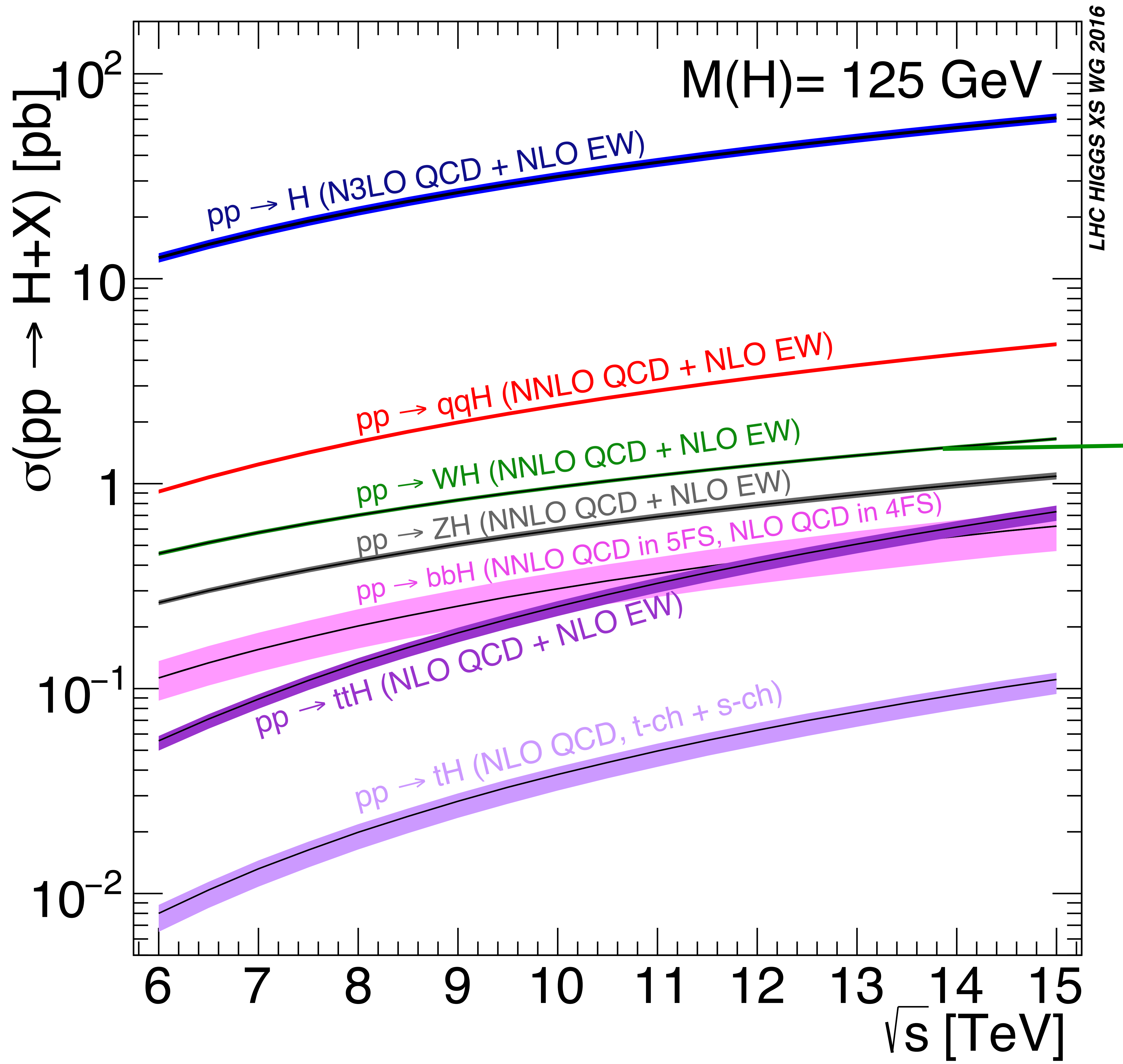


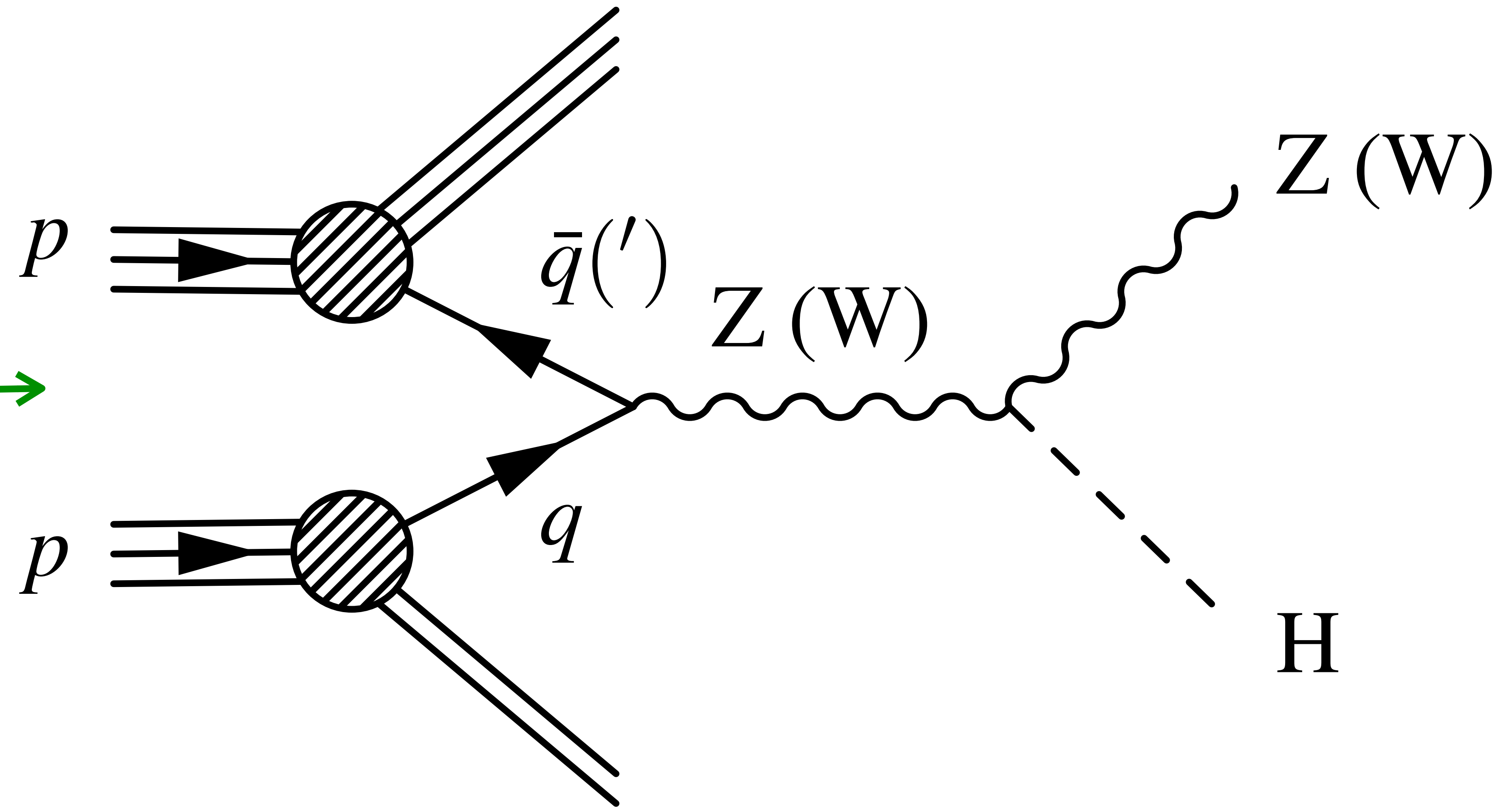
Vector Boson Fusion (VBF) 7%



large background but a very distinctive topology

$H(b\bar{b})$ at LHC

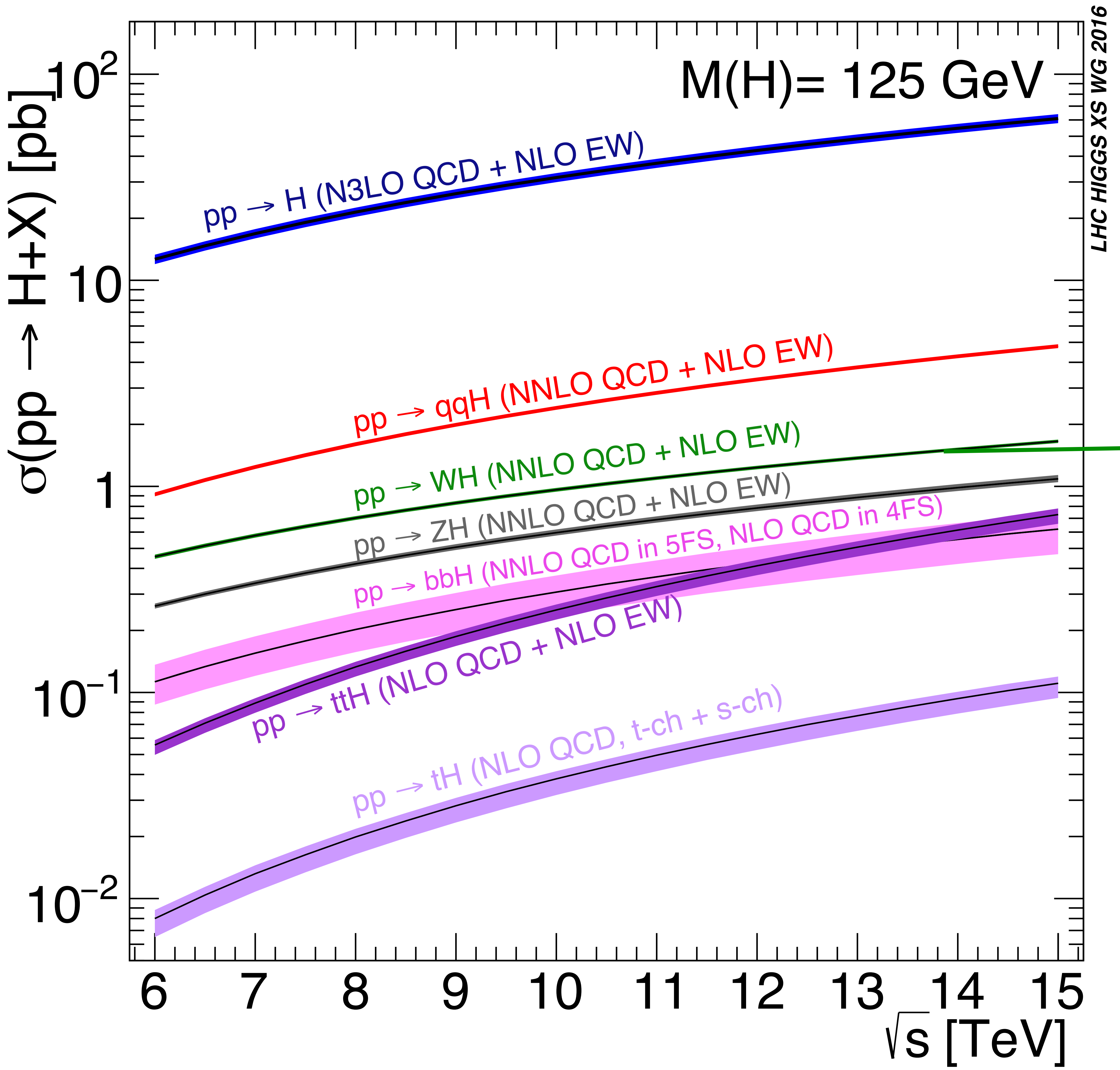




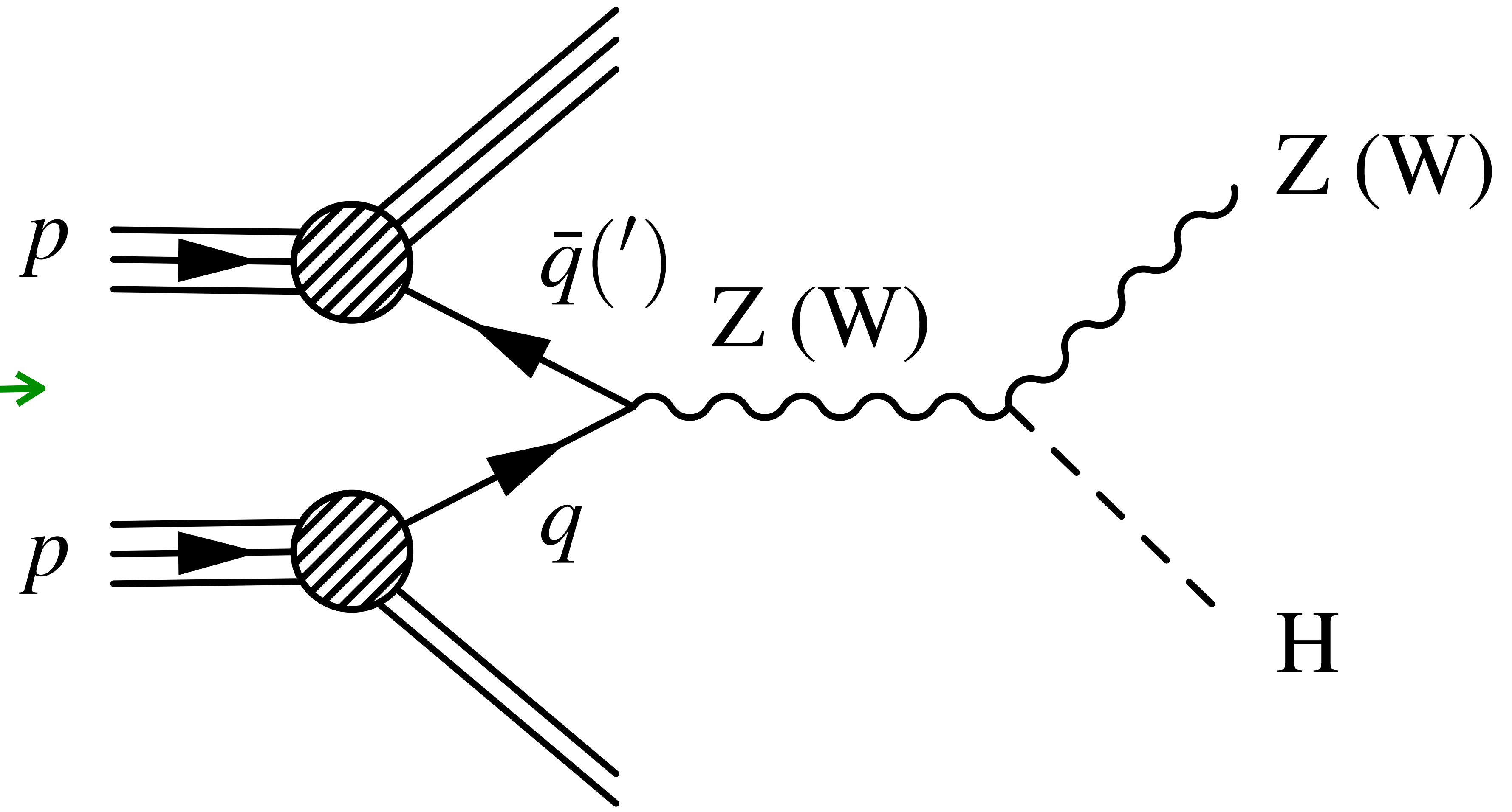
VH 4%

- W/Z decaying leptonically
  - leptons, E_T^{mis} to trigger and suppress backgrounds
- Requiring high p_T V to reduce multijets background
  - S/B at LHC is 2.5x lower than at Tevatron

H($b\bar{b}$) at LHC



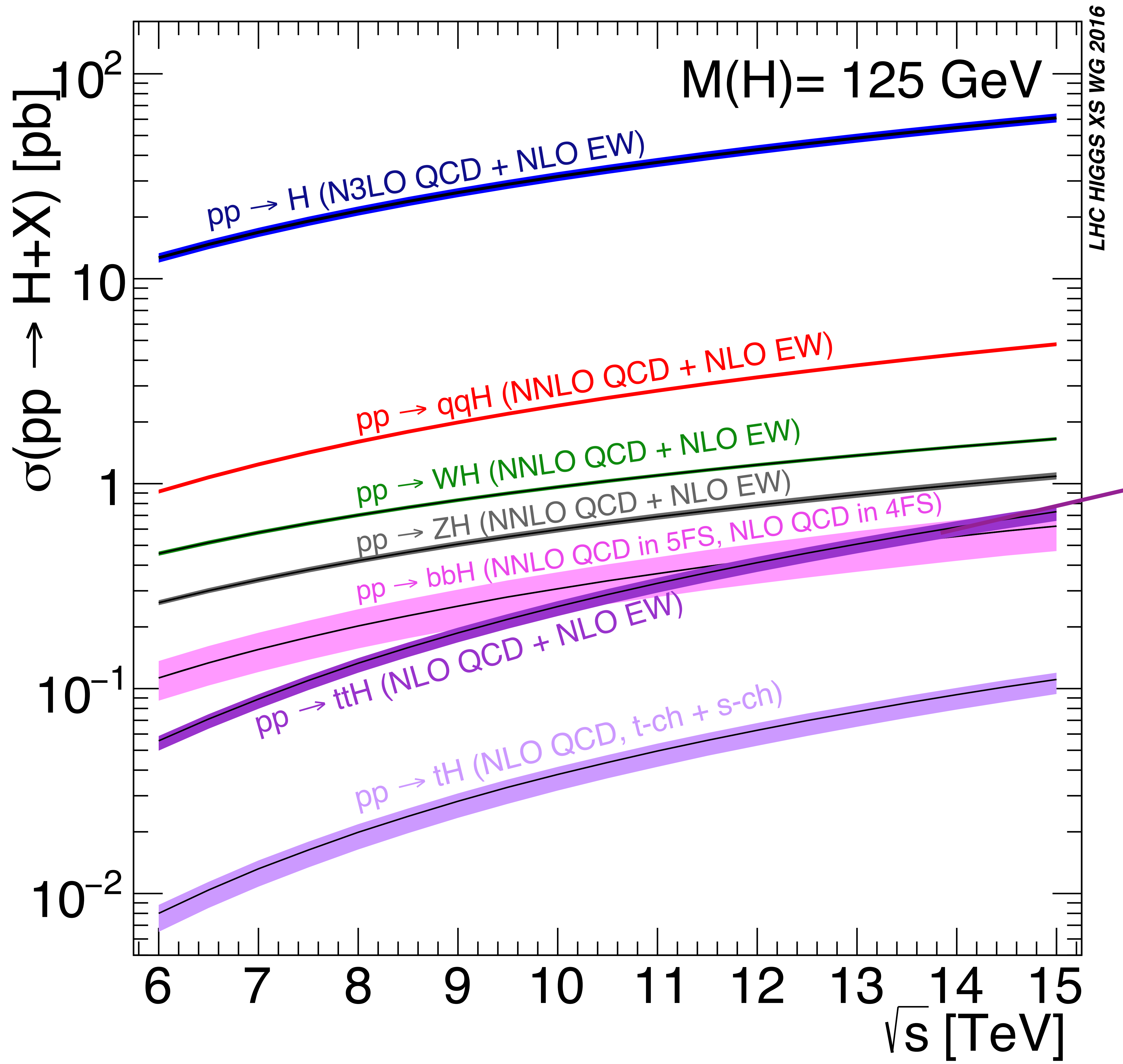
Most sensitive

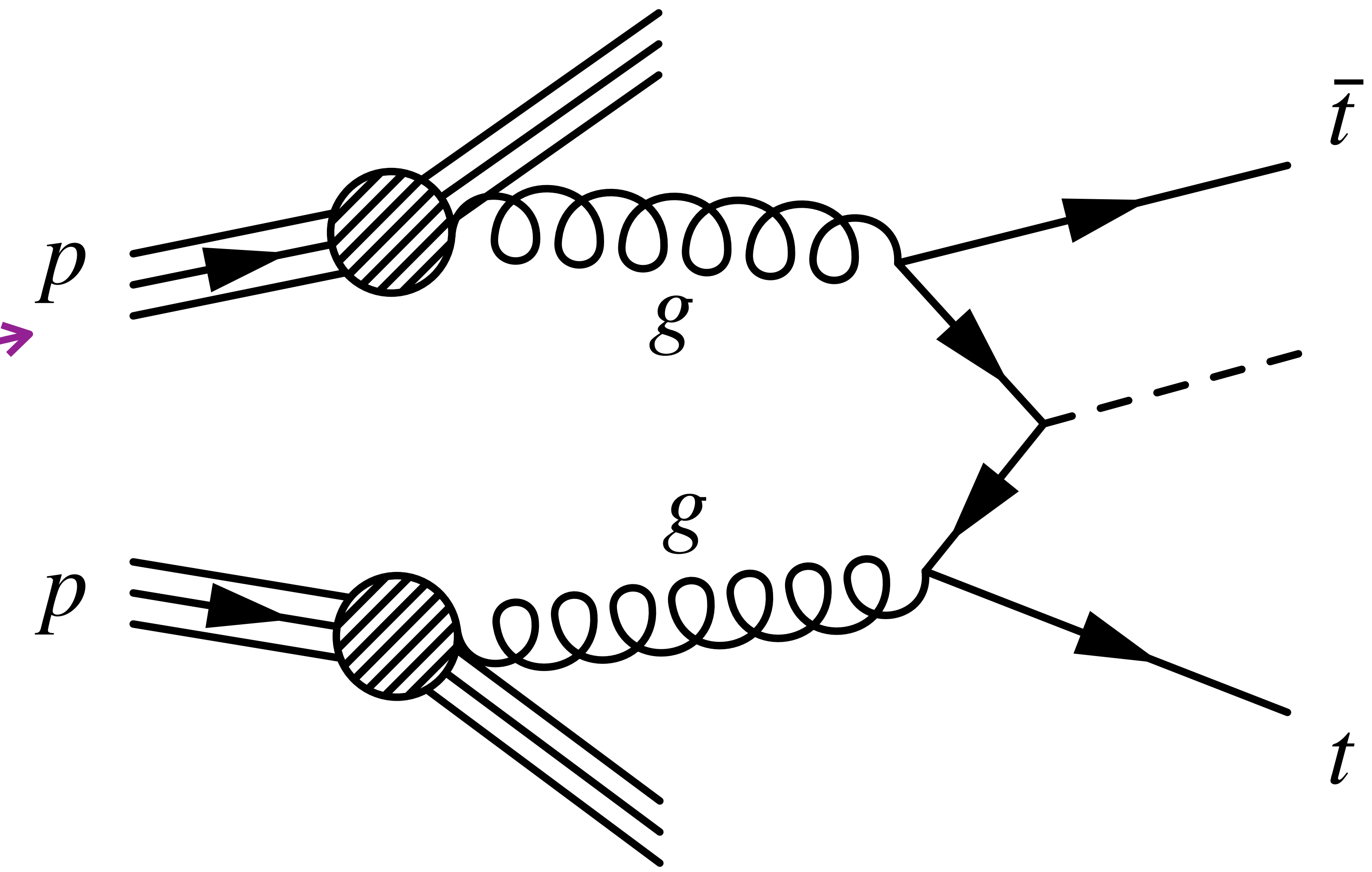


VH 4%

- W/Z decaying leptonically
  - leptons, E_T^{mis} to trigger and suppress backgrounds
- Requiring high p_T V to reduce multijets background
  - S/B at LHC is 2.5x lower than at Tevatron

$H(b\bar{b})$ at LHC





ttH
1%

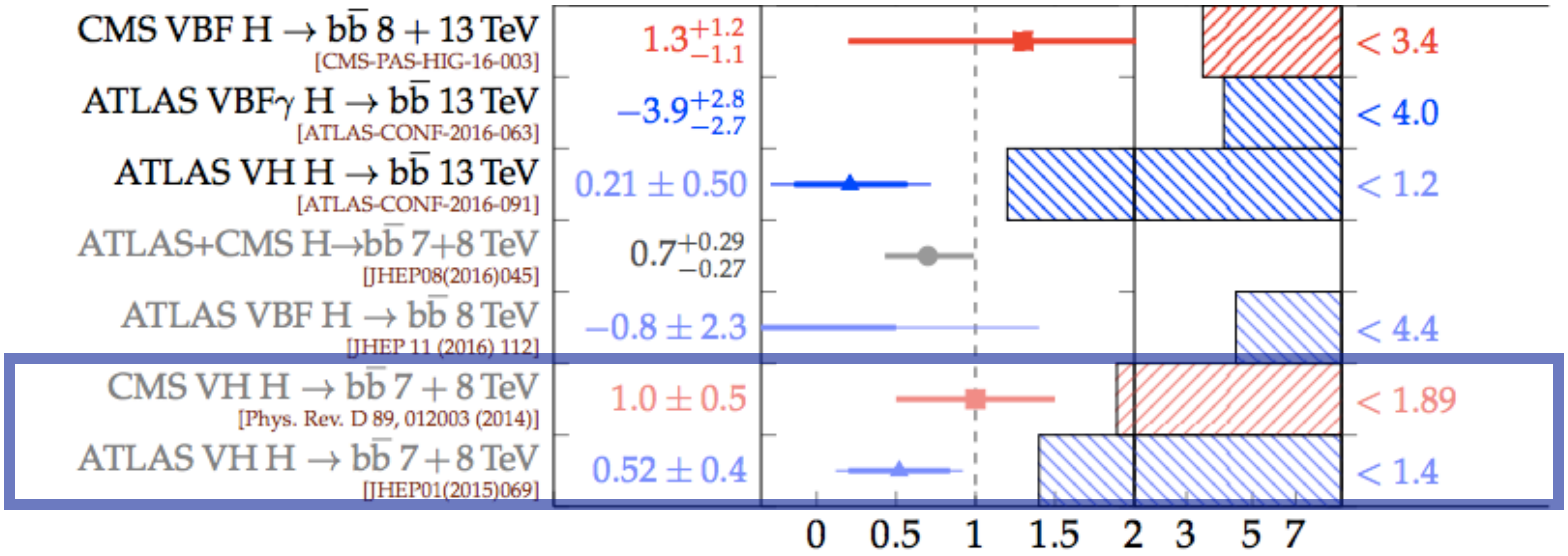
- direct probe of the top-Higgs coupling
- dominant backgrounds is $t\bar{t}$ + jets

Status of H($b\bar{b}$) at LHC

The LHC combination of the Run 1 ATLAS and CMS analyses resulted in a significance of **2.6σ** (3.7σ) **observed** (expected)

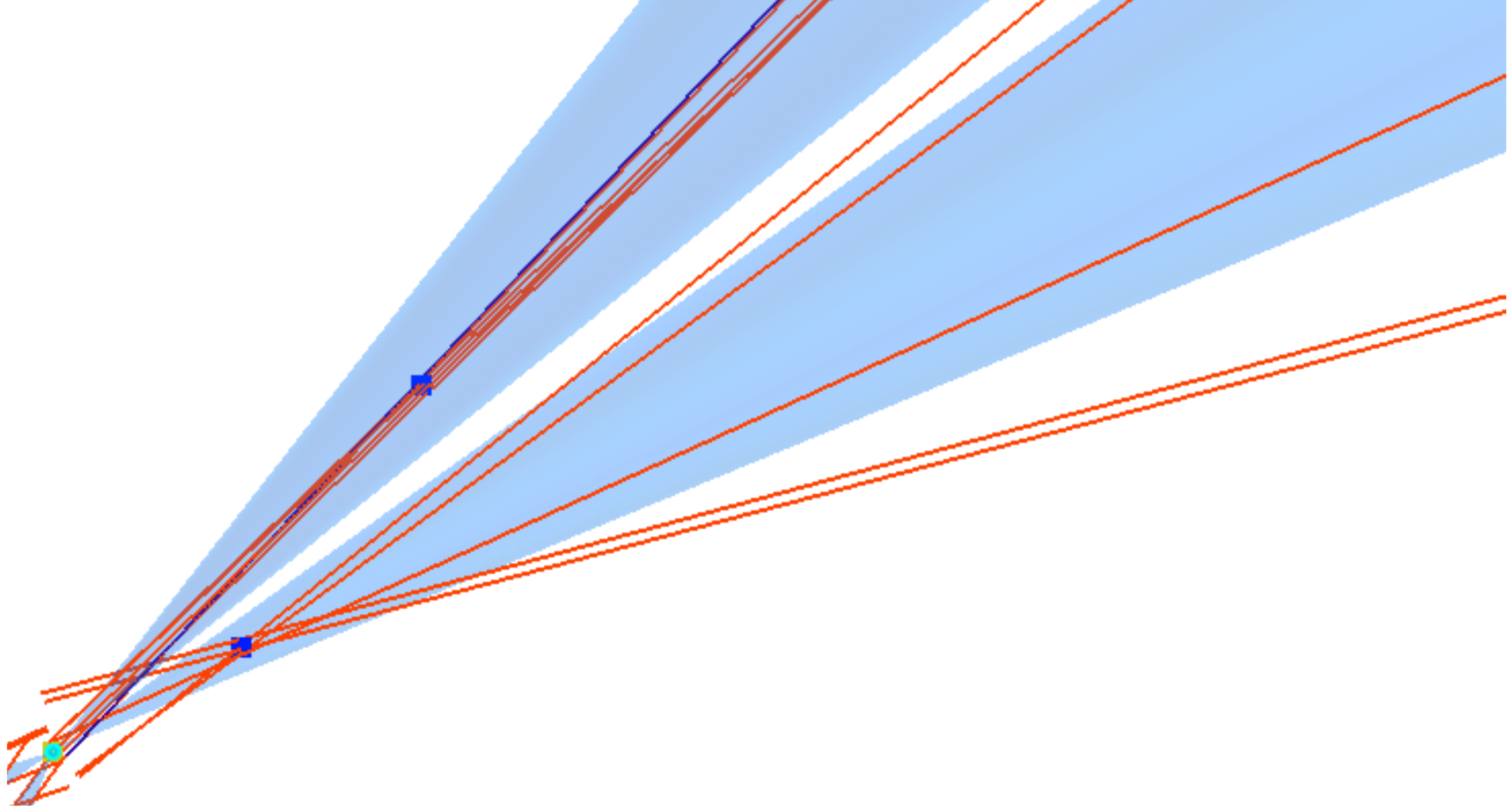
VH is the **most sensitive channel for H($b\bar{b}$)**

courtesy of G. Gaycken



μ = observed signal, in units of the amount predicted from the SM

Outline

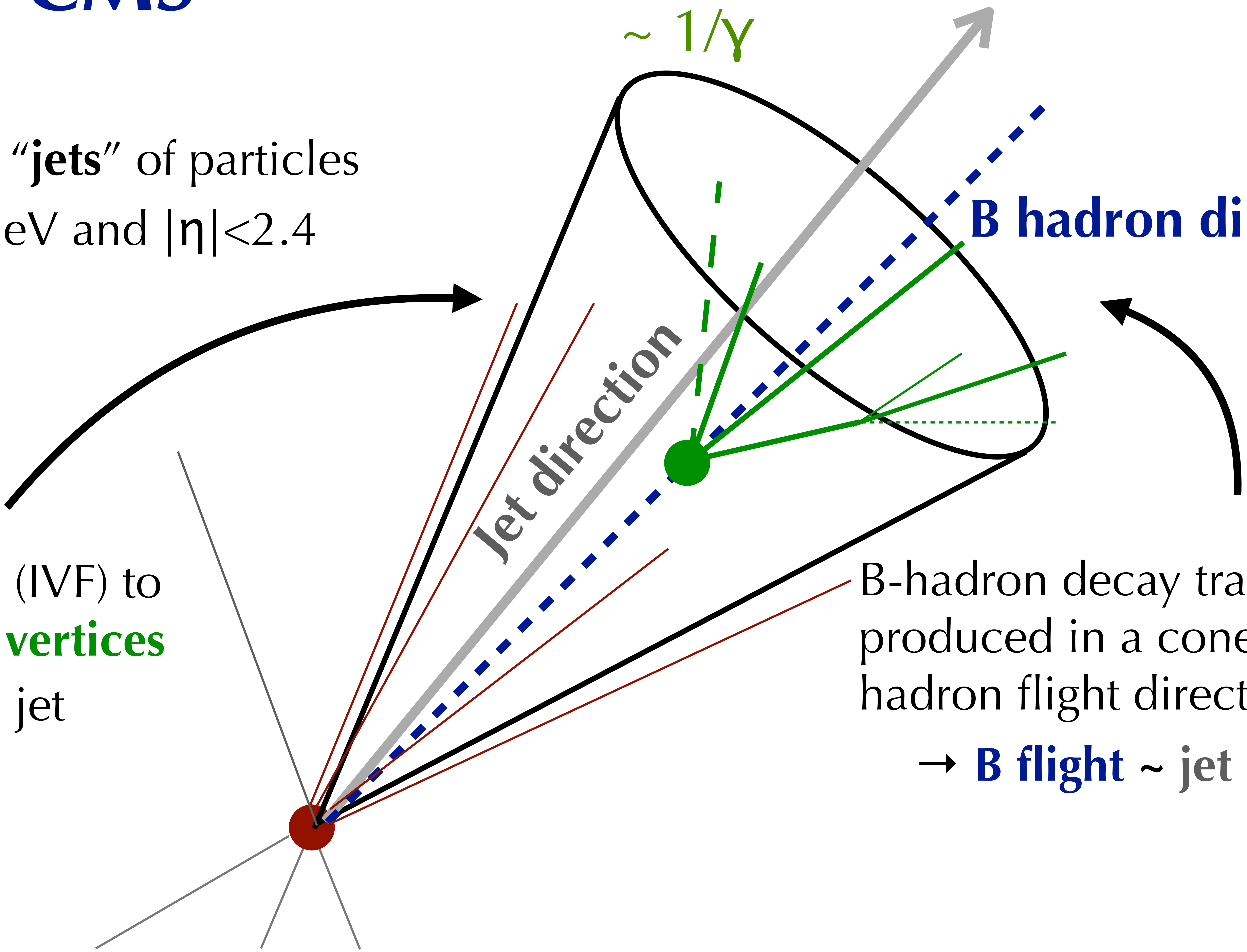


- Higgs boson discovery at LHC
- $H \rightarrow b\bar{b}$ state of the art
- **Tools for identifying $H \rightarrow b\bar{b}$ at high p_T**
    - **b-quark identification in CMS**
- Inclusive search for boosted $H \rightarrow b\bar{b}$
- Future perspectives

b tagging in CMS

b quarks hadronize in “**jets**” of particles

- Jets with p_T> 30 GeV and $|\eta|$<2.4



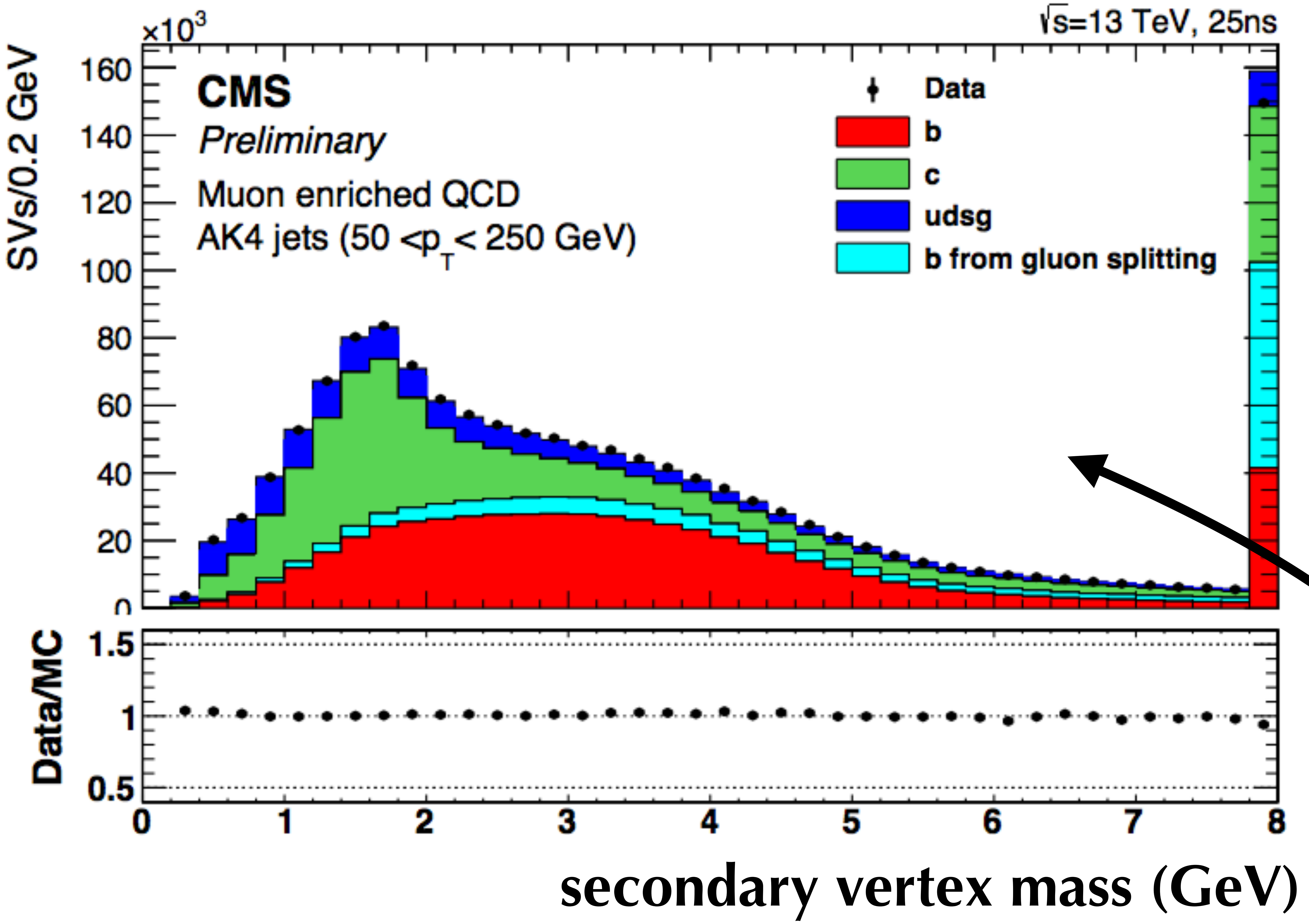
Inclusive Vertex Finder (IVF) to reconstruct **secondary vertices**

- Independent of the jet direction

B-hadron decay tracks are mostly produced in a cone in the B hadron flight direction

→ **B flight ~ jet direction**

b tagging in CMS

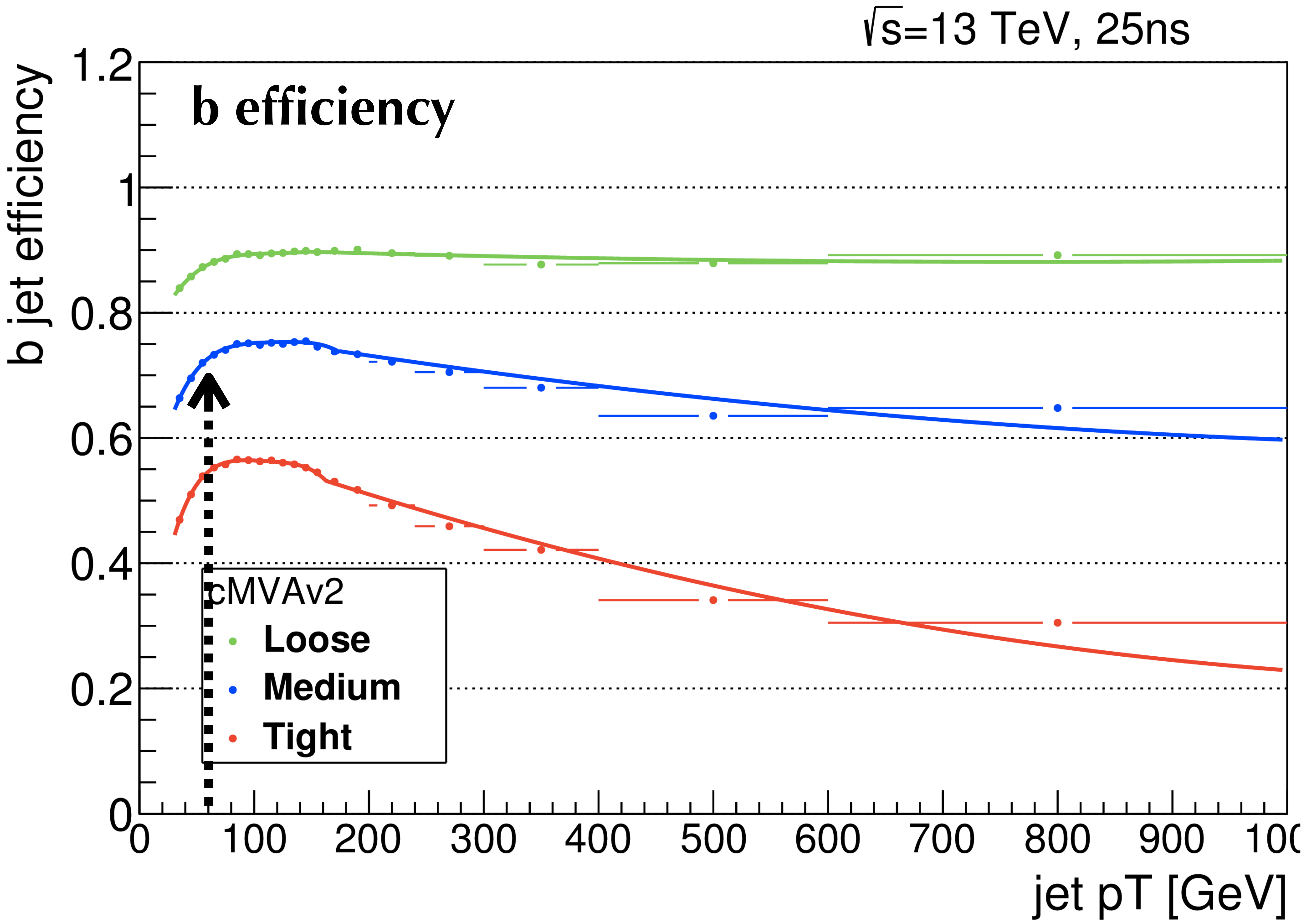


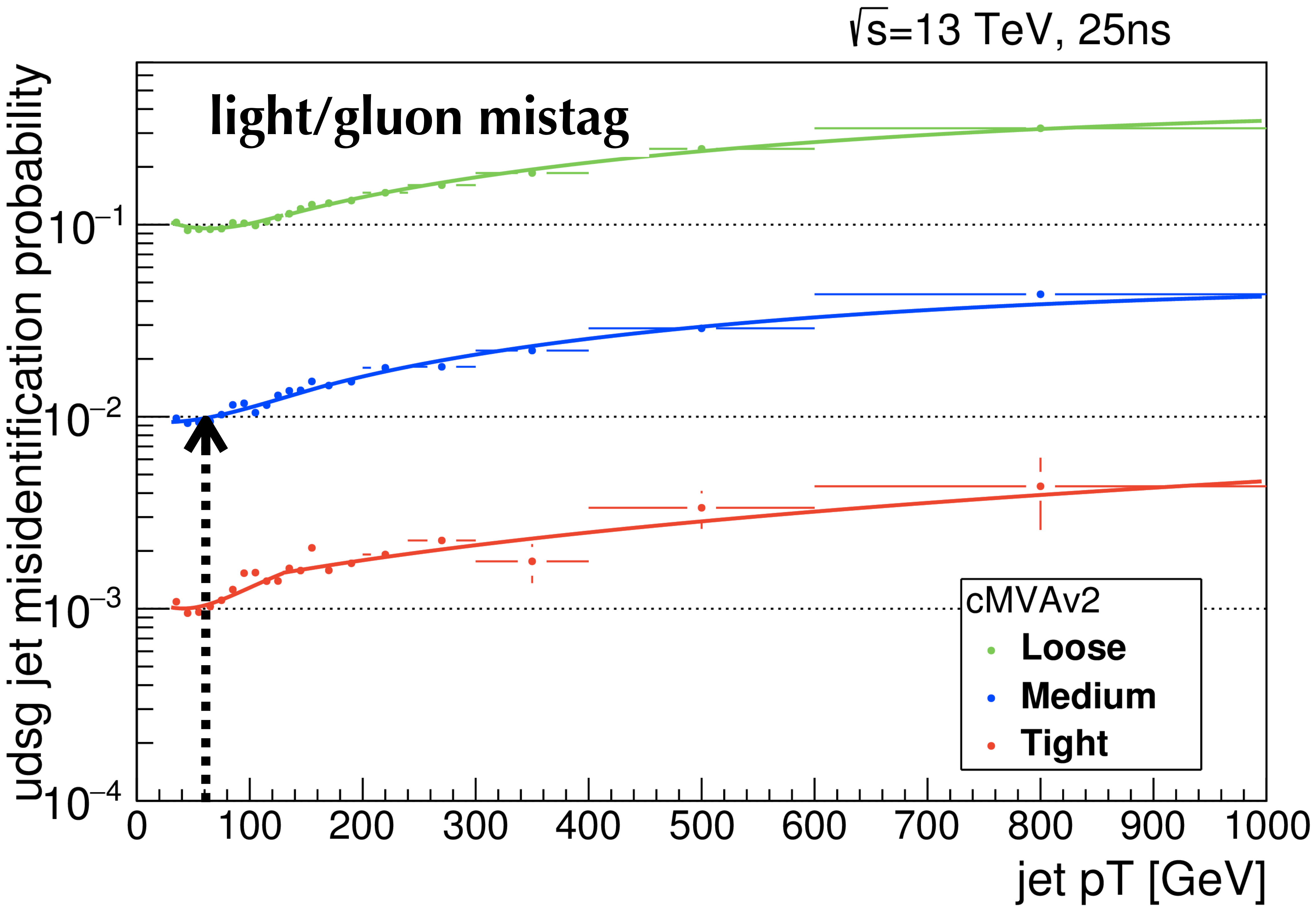
b-tagging algorithms combine with a **multivariate approach** the information from:

- **impact parameter** significance of charged-particle tracks
- the presence of a **lepton** in the jet and its properties
- the presence and properties of reconstructed **secondary vertices**

Performance of b-tagging in CMS

CMS-PAS-BTV-15-001





optimal working point for a **H to $b\bar{b}$ search** search has 70% b efficiency and 1% mistag probability

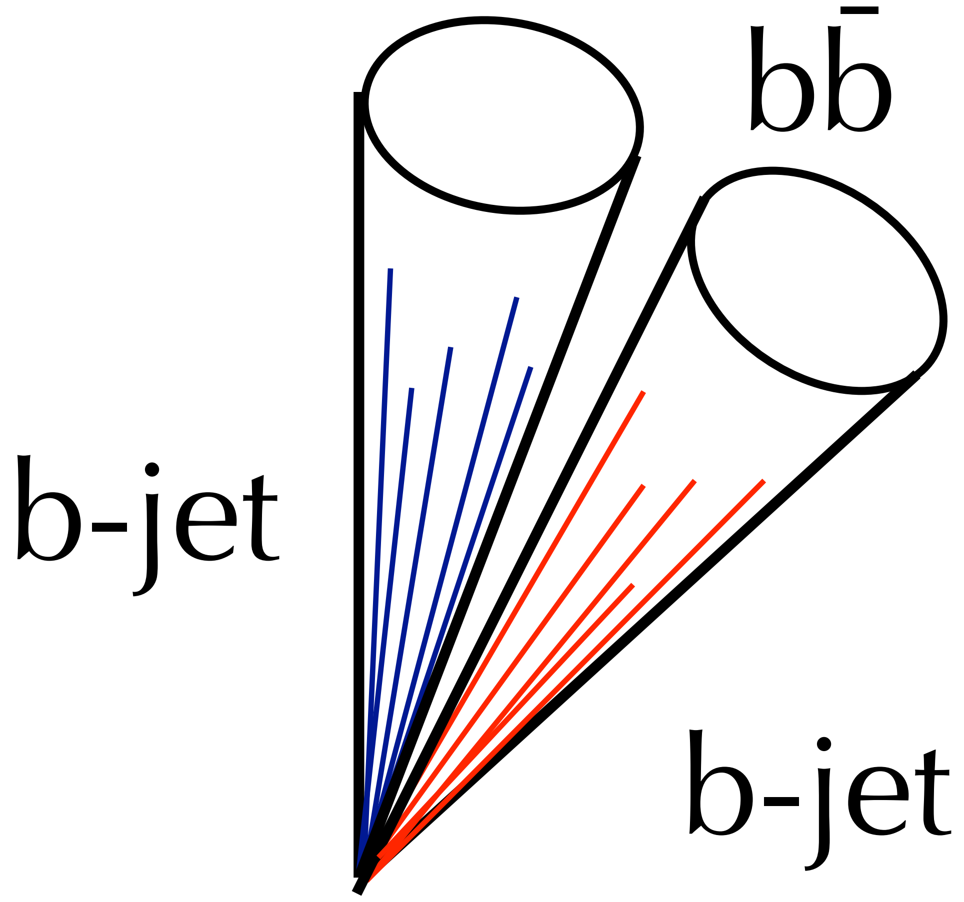
boosted H($b\bar{b}$)

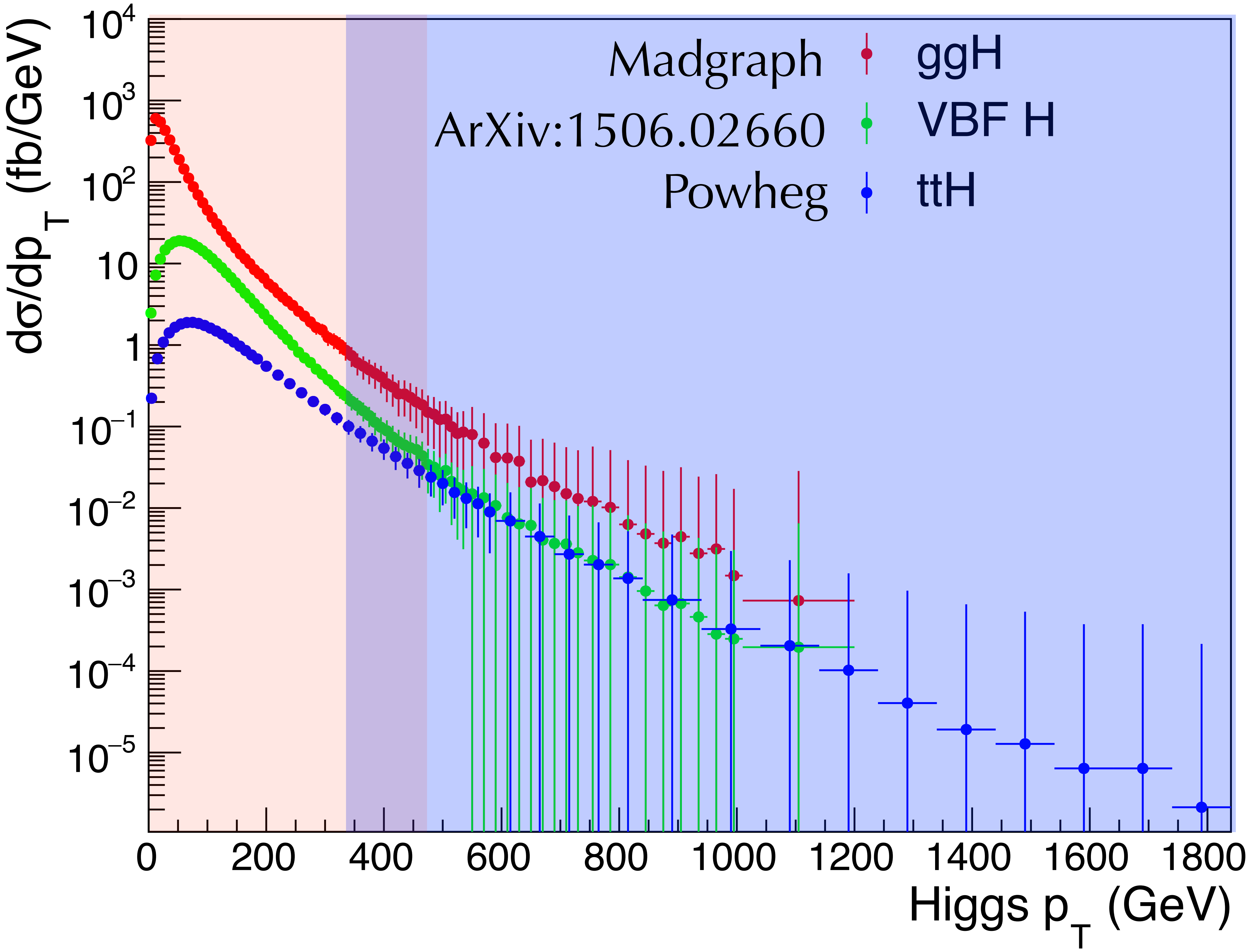
$dR(b\bar{b}) \sim 2m_H/p_T$

300 400 500 600

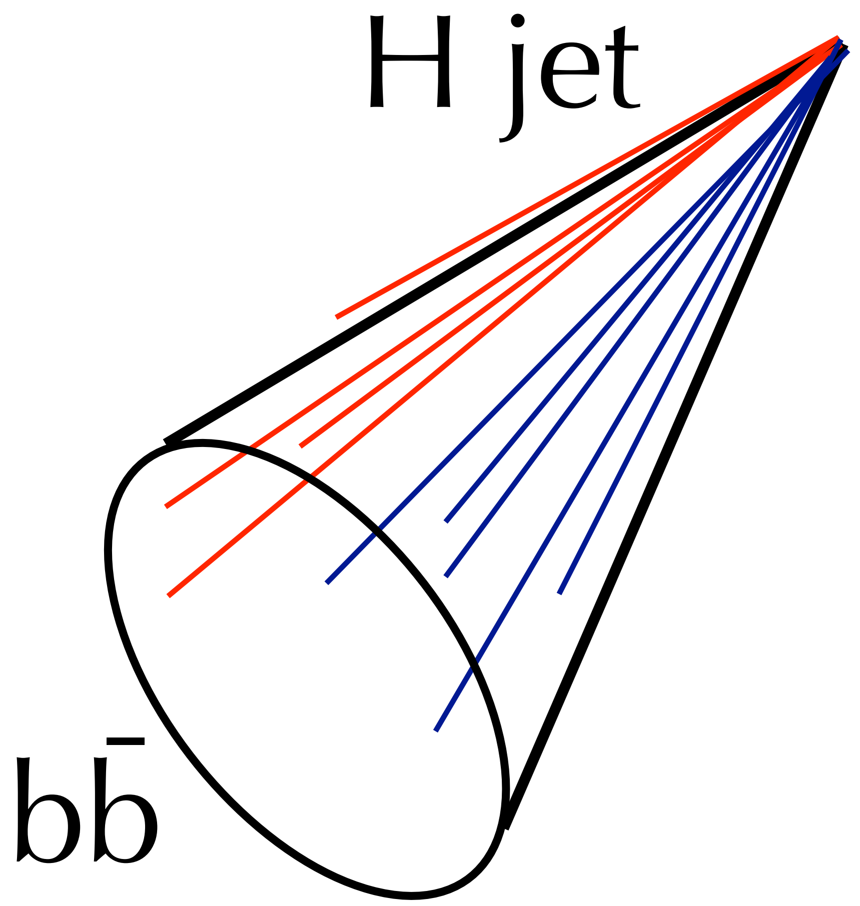
H p_T (GeV)

two-separate b-jets (R = 0.4)

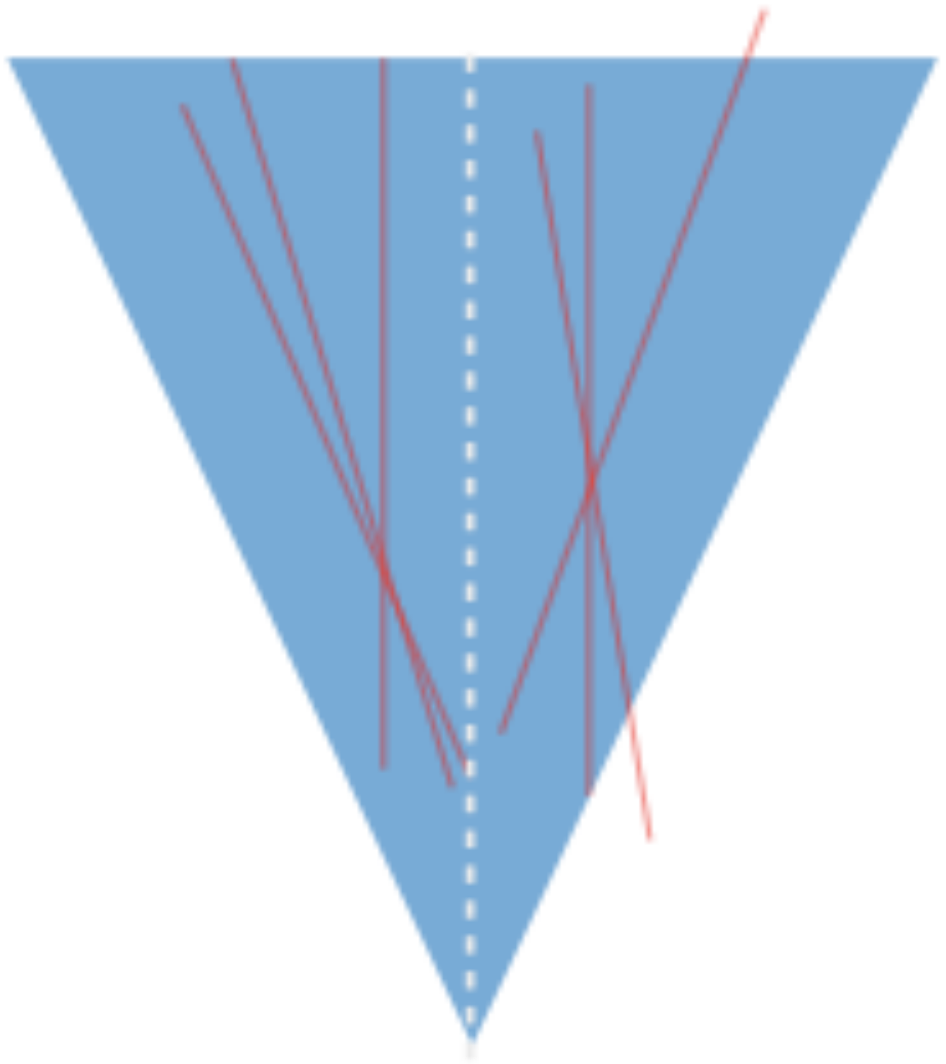




one single large-cone (fat) jet (R = 0.8)

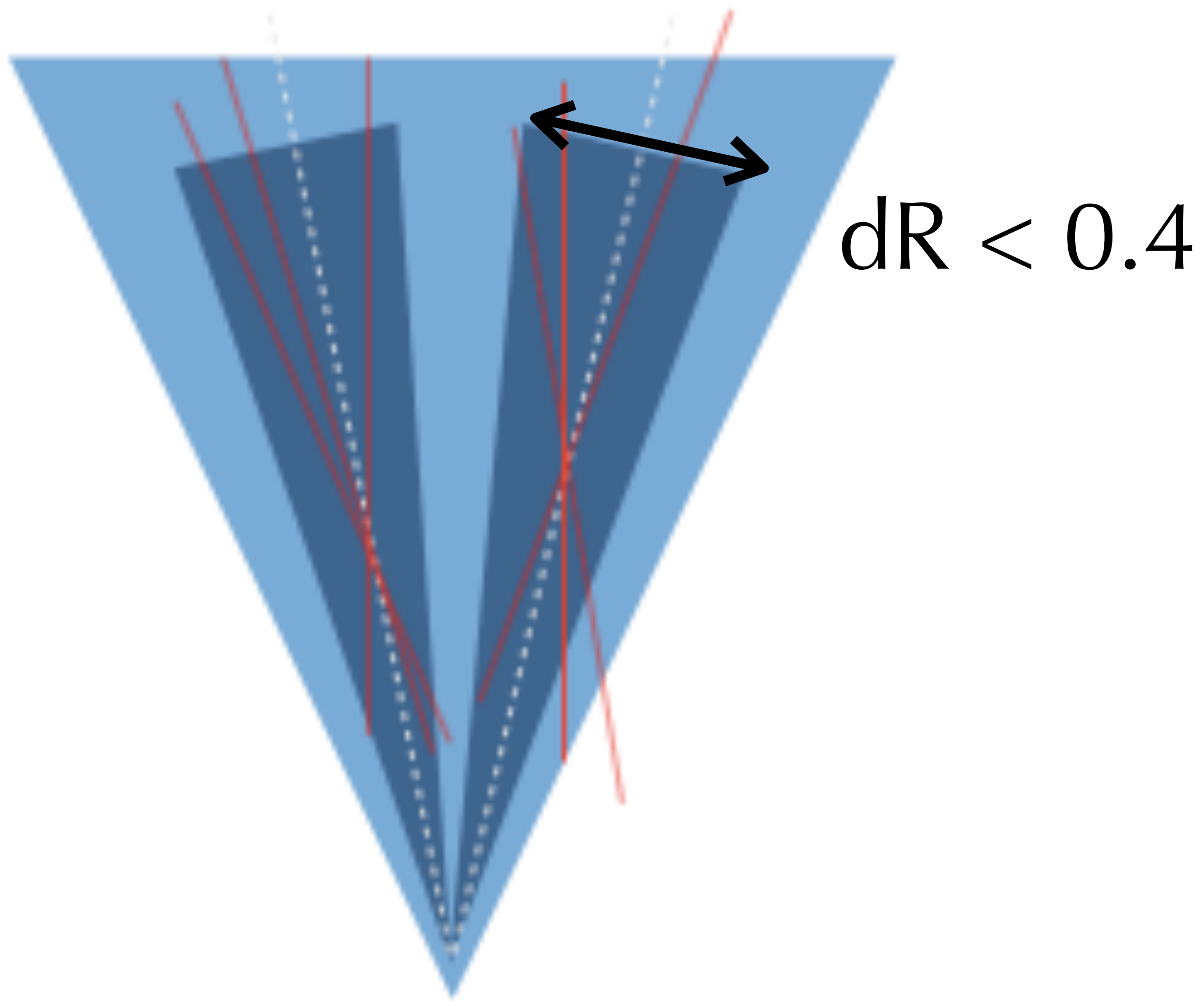


b-tagging, multiple approaches



fat-jet b-tagging

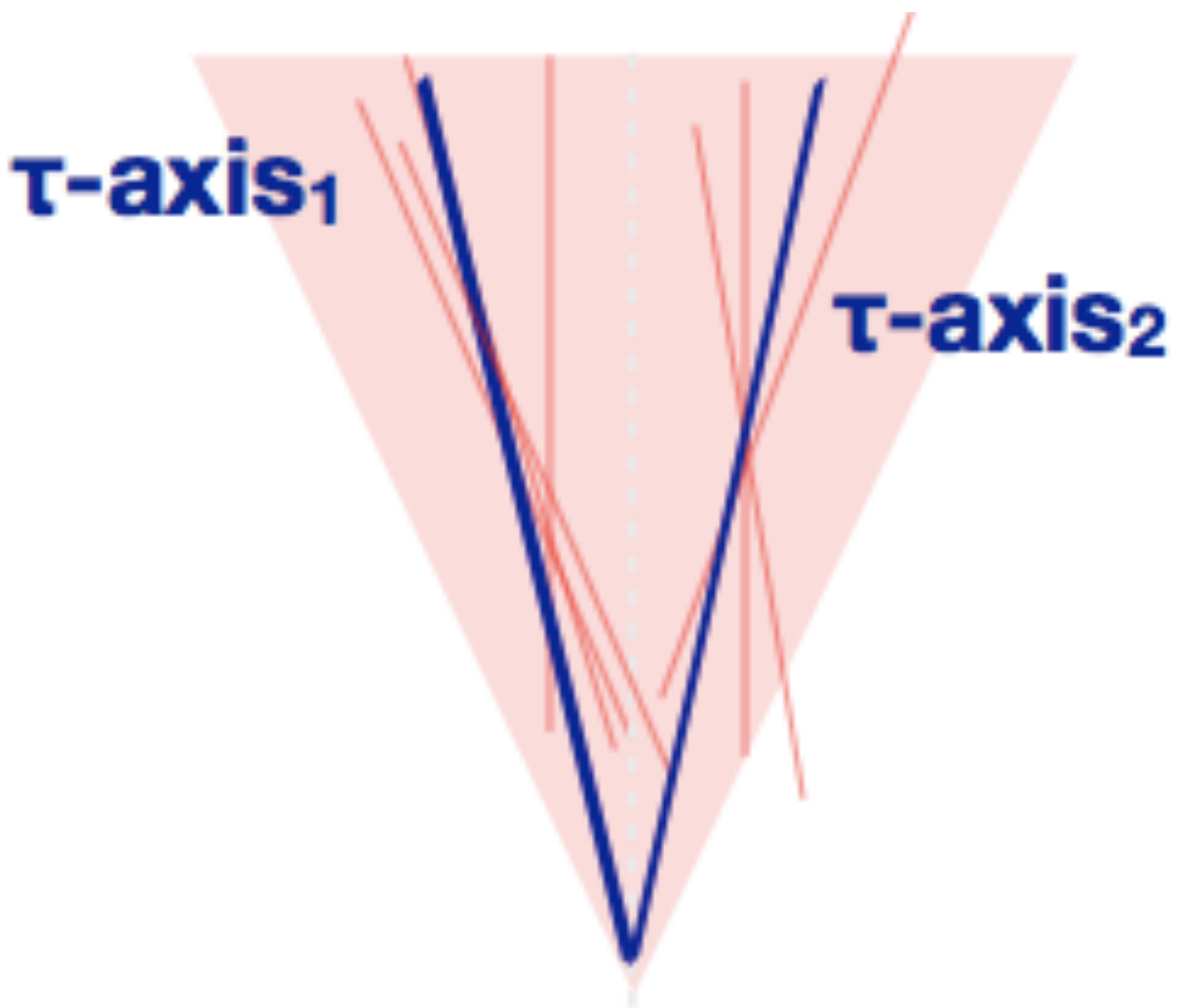
- Based on the standard b-tagging algorithm
- Not designed for tagging two b's in the same jet



sub-jet b-tagging

- Defines sub-jets
- Standard b-tagging algorithm applied to each subjet

NEW

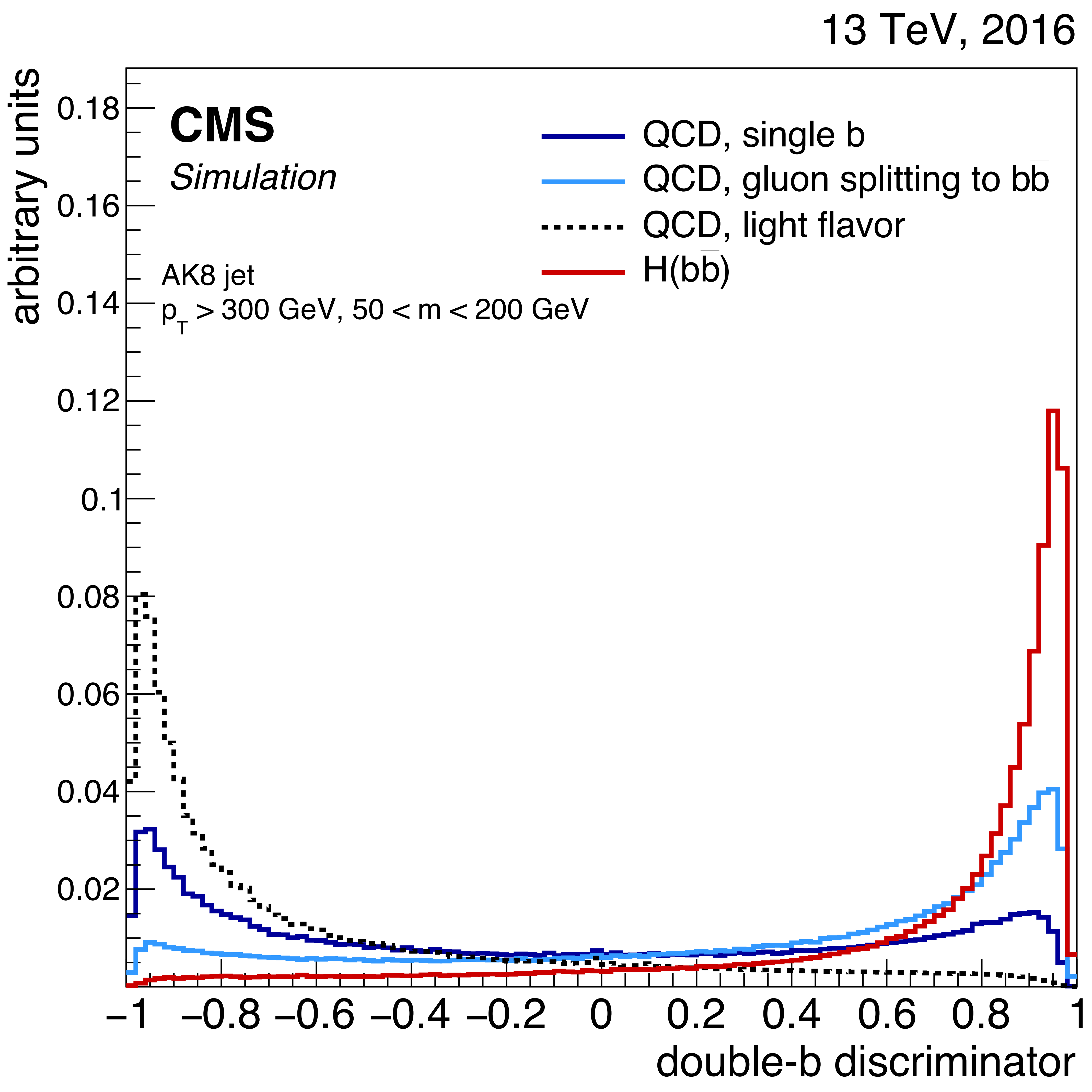


double-b tagger

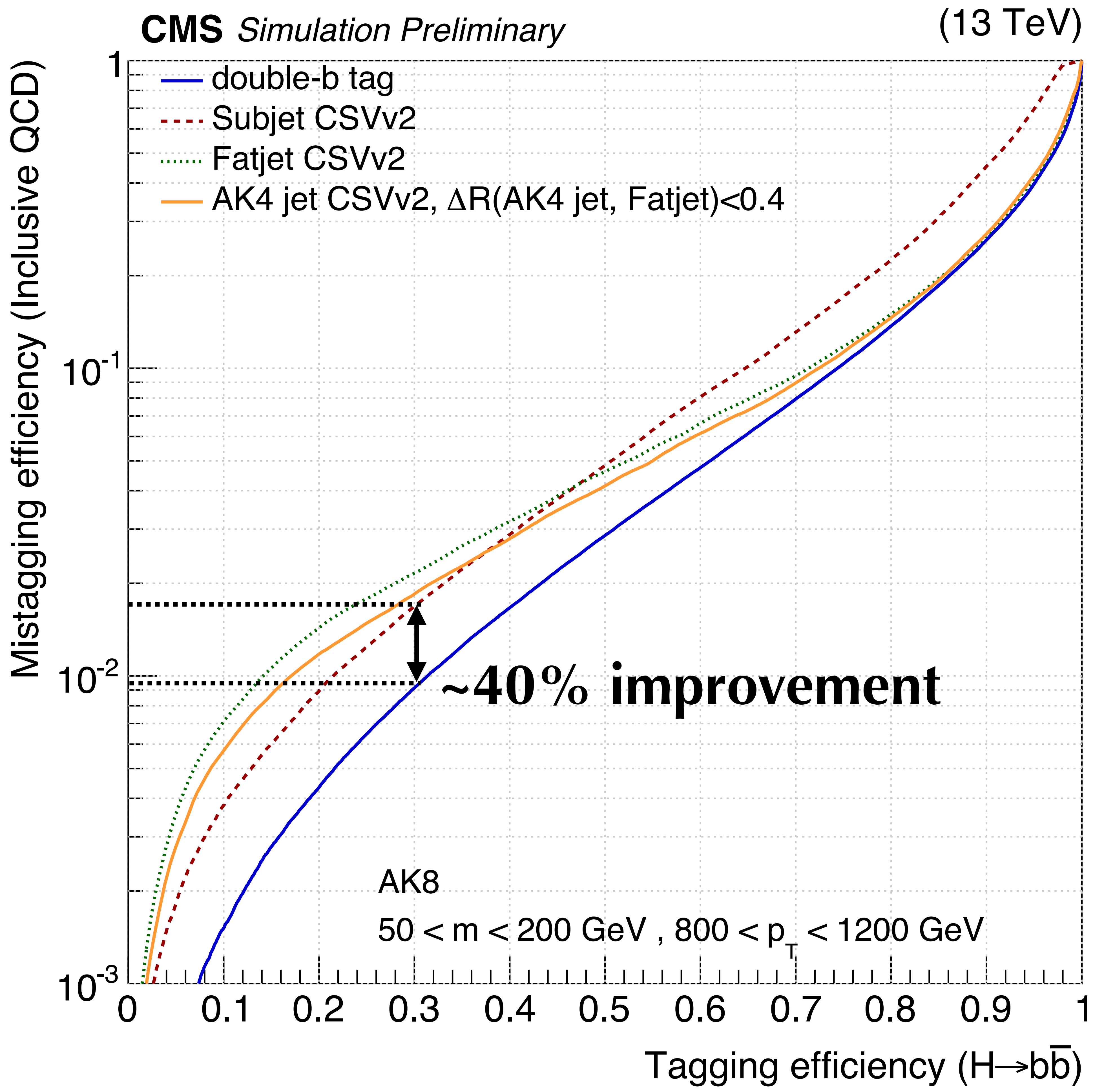
- Identifies the two B hadron decay chains from b and $\bar{b}$ within the same fat jet.
- It does not define sub-jet but uses N-jettiness axes

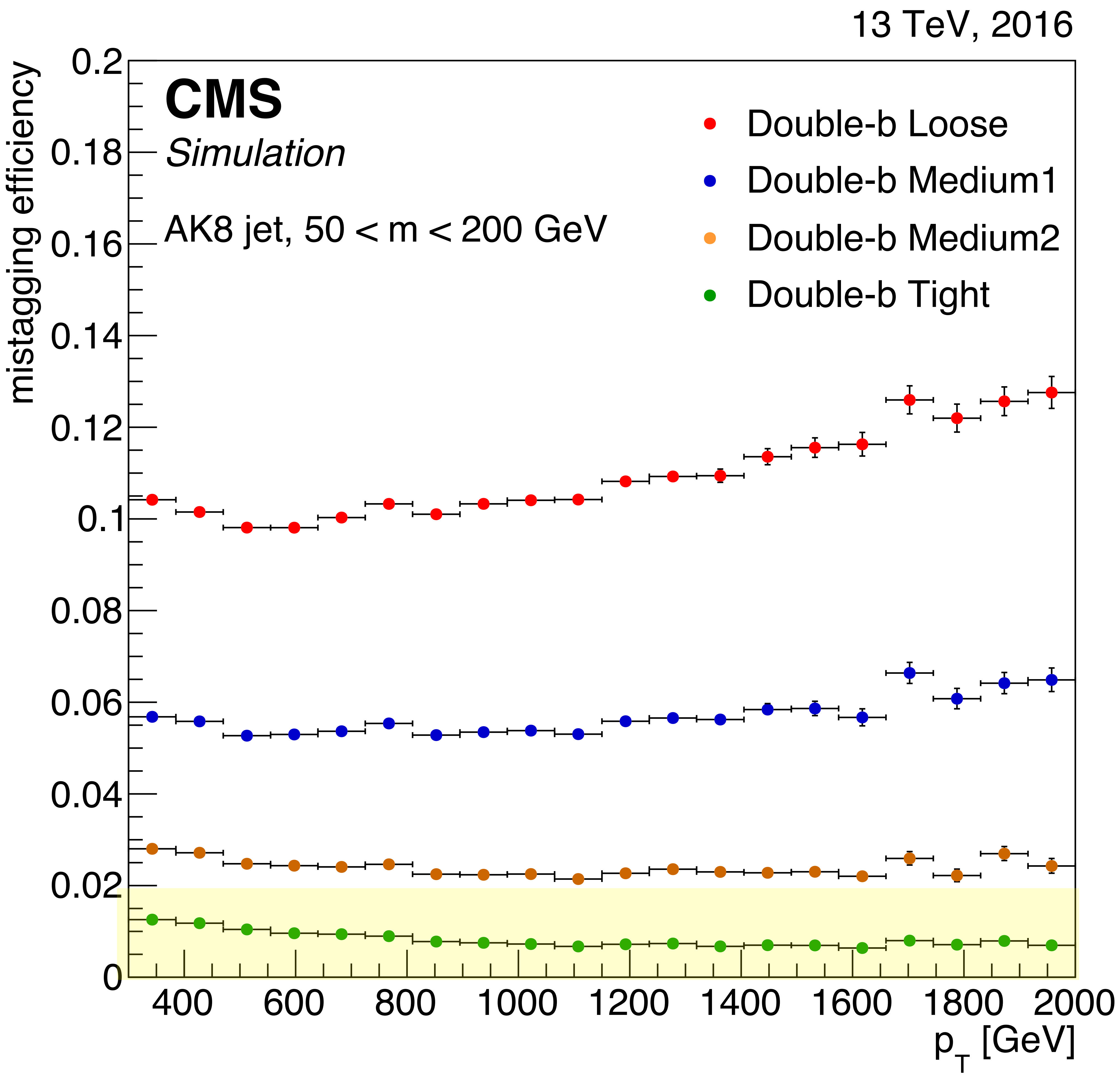
double-b tagger

- Combines **tracking and vertexing information** with a multivariate approach
  - 27 observables are used
- It **targets the $b\bar{b}$** signal aiming to be:
  - ***mass independent***
  - ***p_T independent***
    - **training strategy** is designed to cover a very wide p_T range
    - inputs are chosen to avoid p_T correlation
      - *no dR-like variables, no substructure info*



Efficiency vs. Mistag rate



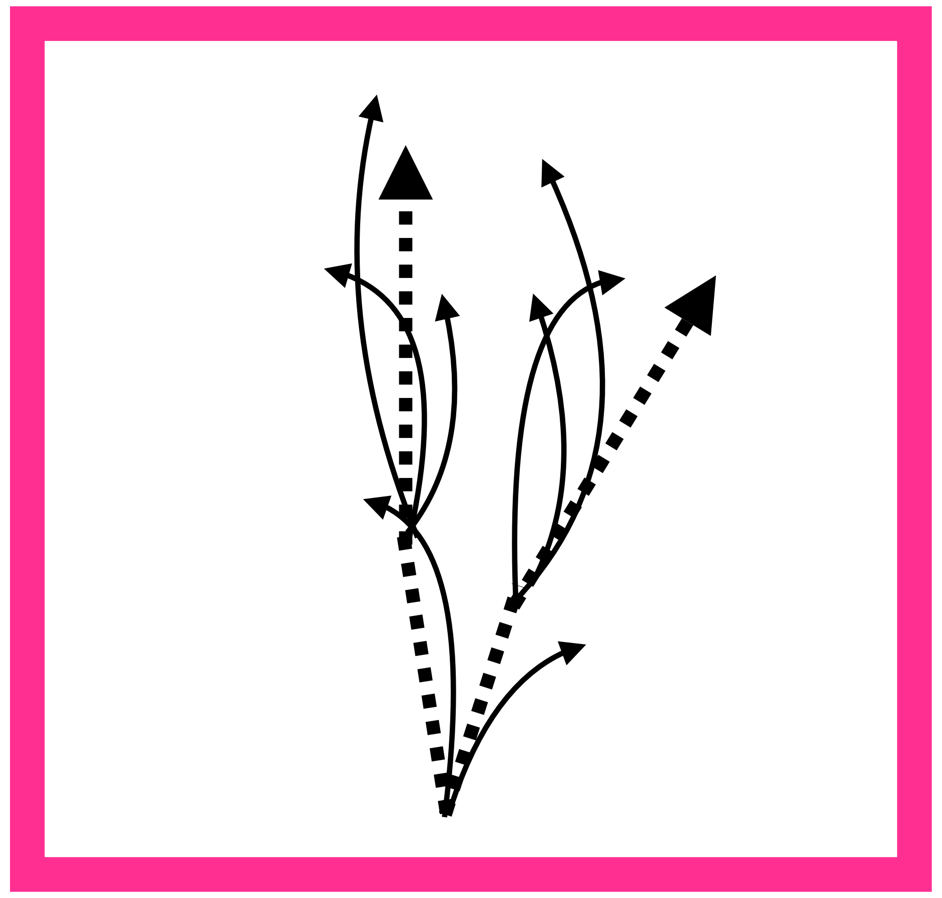


The mistag rate is approximately flat across the p_T range by design
Critical point for searches (background estimate)

H-tagging

The boosted $H(b\bar{b})$ signal is identified as large cone size jets:

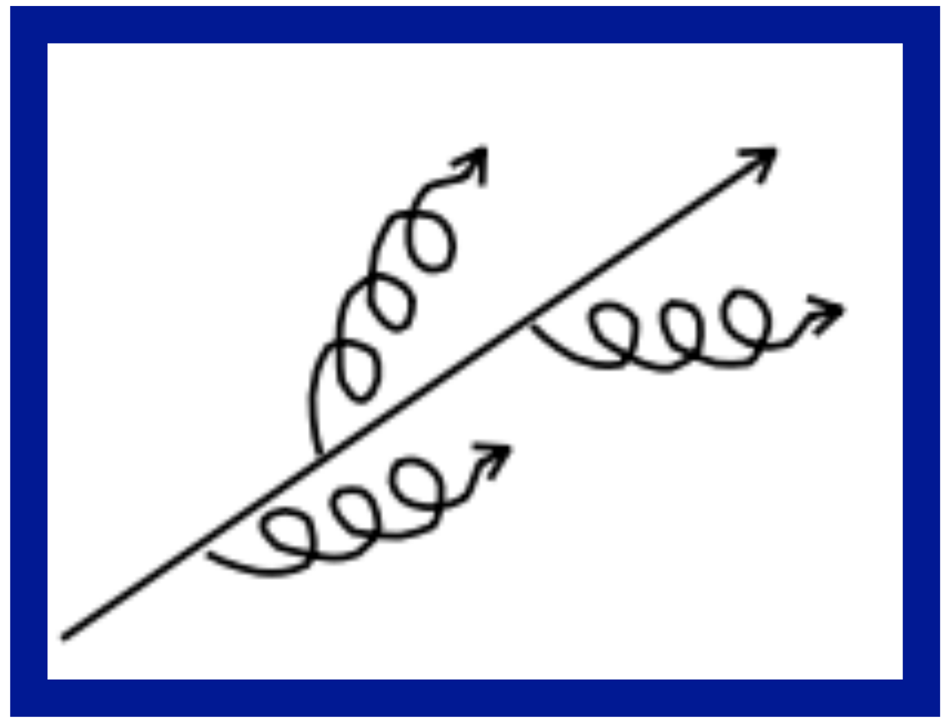
- **R=0.8**
- **PUPPI** (PileUp Per Particle Id) is used to mitigate **pile up effects**



$H/Z(b\bar{b})$

Our tools:

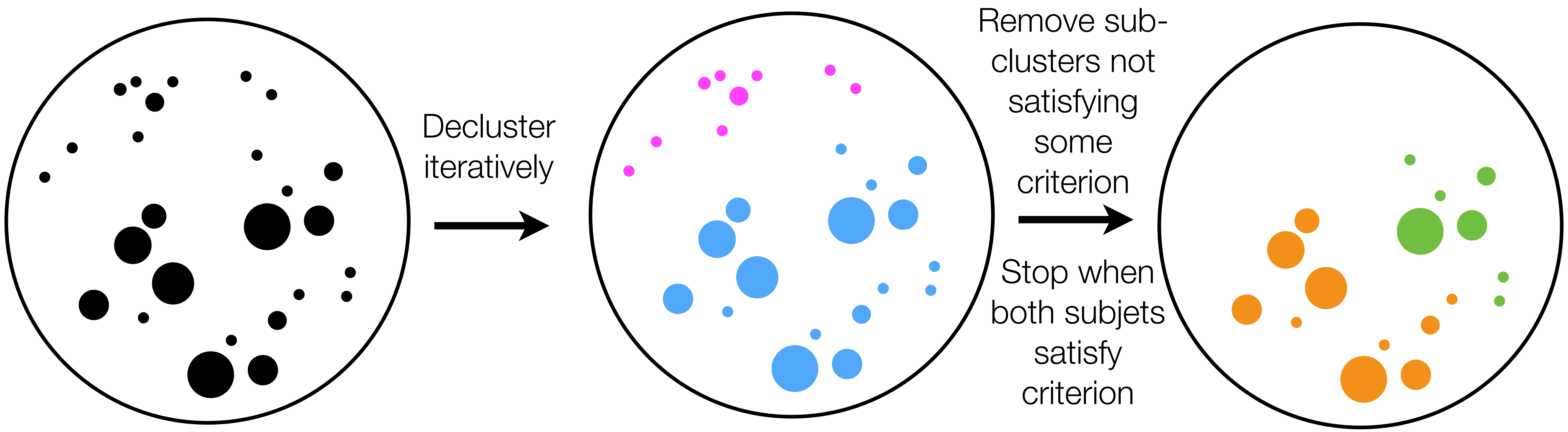
- **b-tagging** to reconstruct the two B hadrons from the b and $\bar{b}$ within the same fat jet
- jet **mass** compatibility with the Higgs
- the composite nature of the jet using **substructure**



background q/g

Jet mass

- Provides good separation between W/Z/H-jets from q/g jets
- **Grooming** removes soft and wide-angle radiation (**soft drop/modified mass soft drop**)

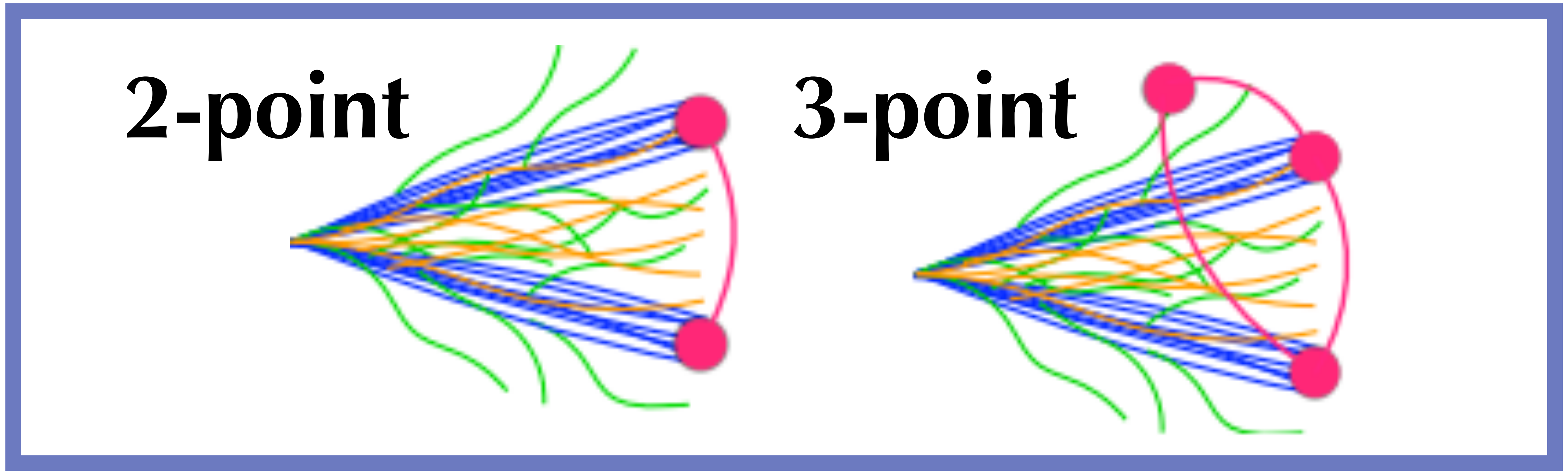


$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\mathrm{cut}} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta} \qquad \beta = 0,\ z_{\mathrm{cut}} = 0.1$$

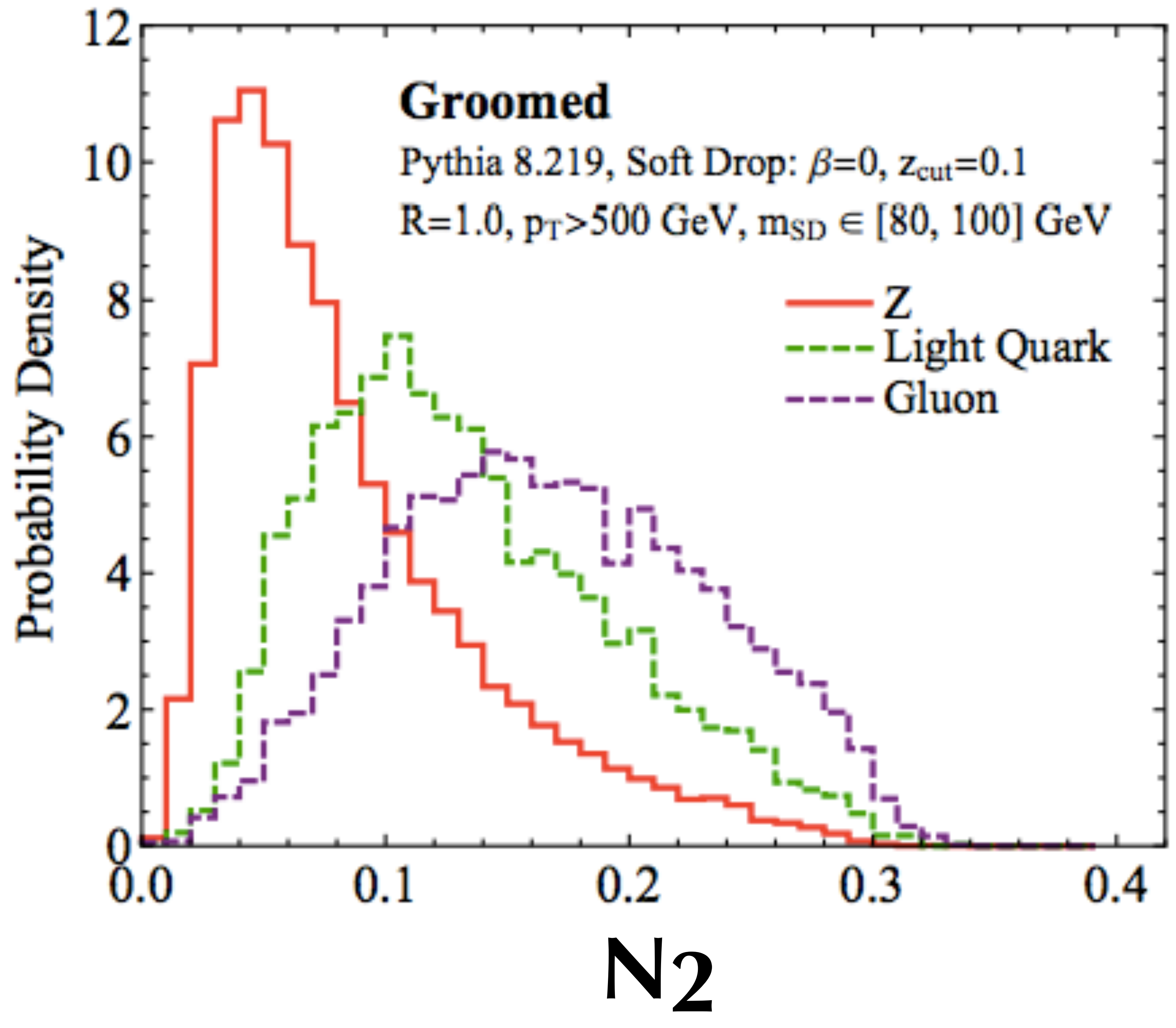
Jet Substructure

ArXiv:1609.07483

- Measures the degree to which a jet can be considered as composed of N prongs
- **Energy correlation functions** are sensitive to N-point correlations in a jet
  - A 2-pronged jet will have $e_3<e_2$

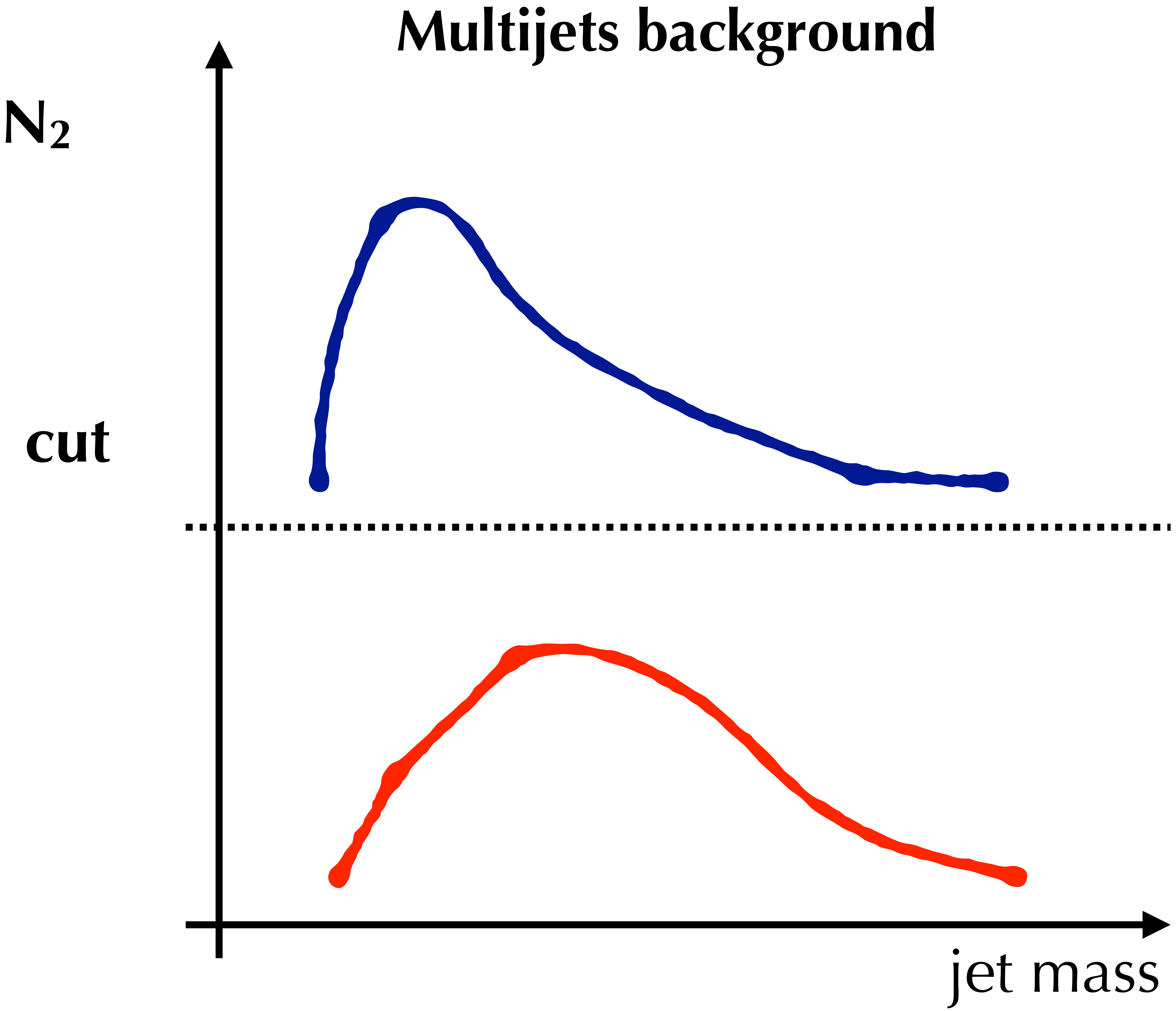


$$N_2 = \frac{e_3}{(e_2)^2}$$



Jet Substructure

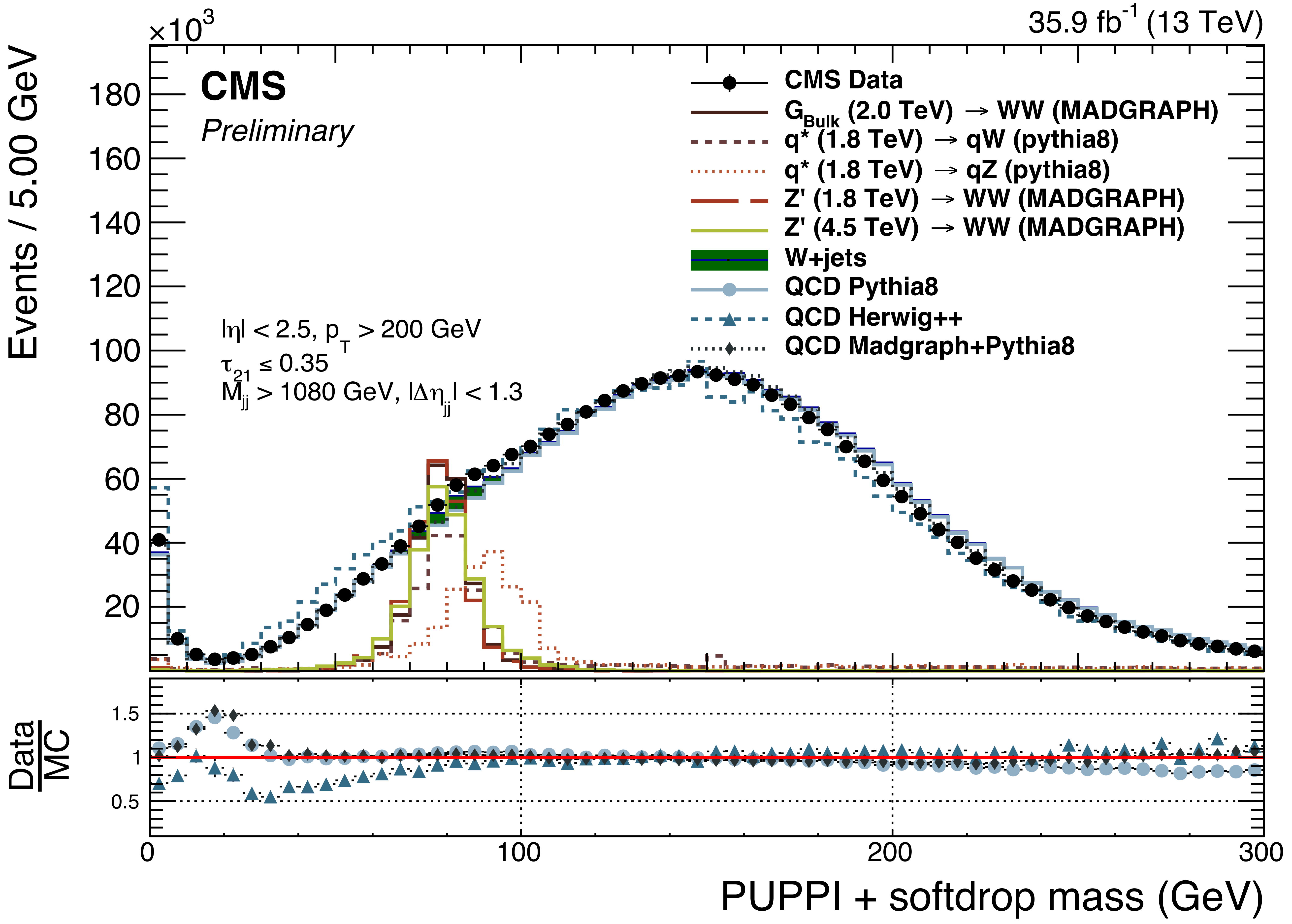
ArXiv:1603.00027
CMS-PAS-B2G-17-001
CMS-PAS-EXO-17-001



N_2 sculpts the jet mass distribution

Jet Substructure

ArXiv:1603.00027
CMS-PAS-B2G-17-001
CMS-PAS-EXO-17-001



N_2 sculpts the jet mass distribution

Jet Substructure

ArXiv:1603.00027
CMS-PAS-B2G-17-001
CMS-PAS-EXO-17-001

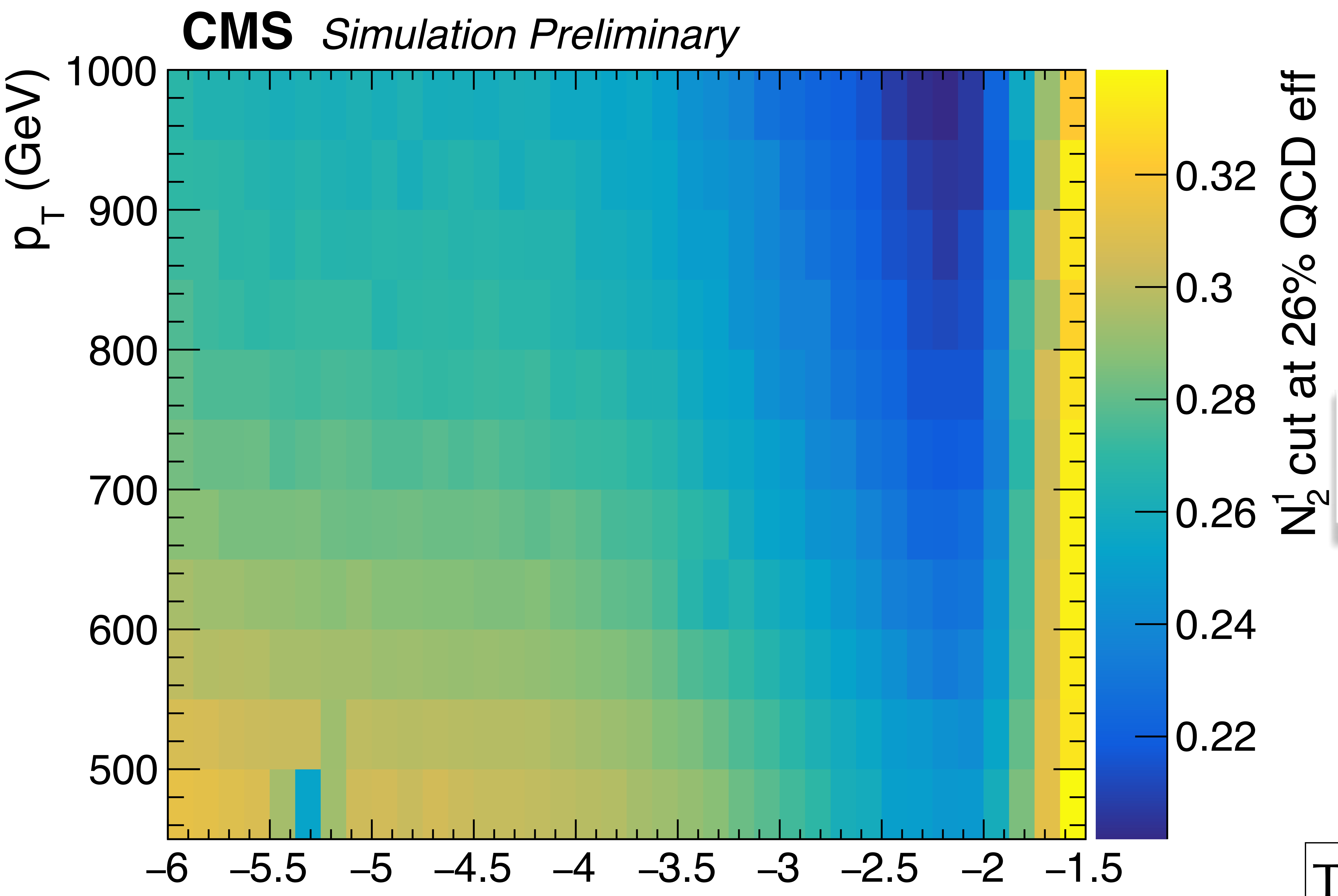
N_2 sculpts the jet mass distribution

We use a **mass-decorrelated** version for the background

$$\mathrm{N}_2^{\mathrm{DDT}} = \mathrm{N}_2 - \mathrm{N}_2(\text{cut at } 26\% \text{ QCD eff.})$$

Jet Substructure

ArXiv:1603.00027
CMS-PAS-B2G-17-001
CMS-PAS-EXO-17-001



$\rho = \log(m_{SD}^2/p_T^{\,2})$

N_2 sculpts the jet mass distribution

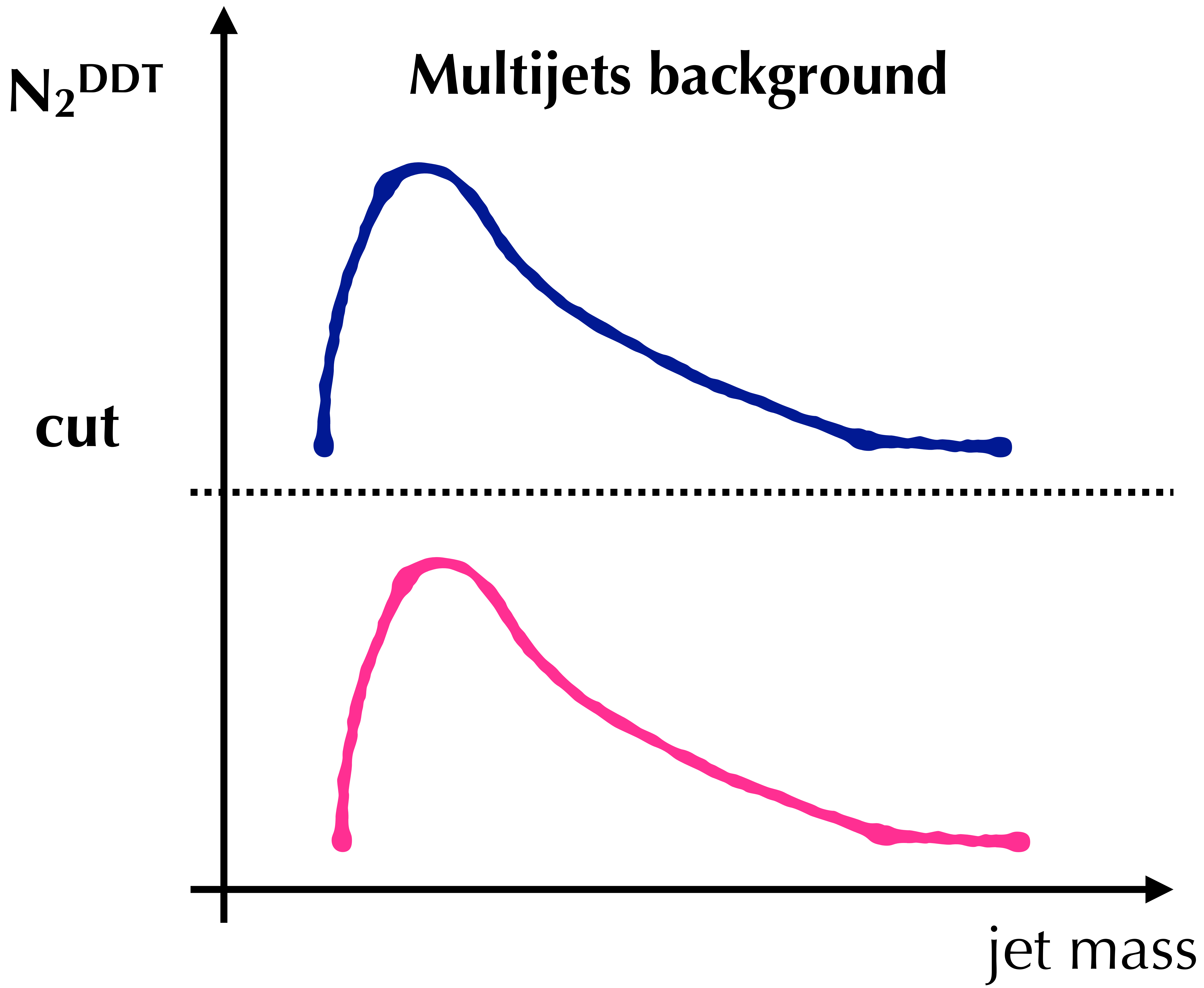
We use a **mass-decorrelated** version for the background

$$N_2^{DDT} = N_2 - N_2(\text{cut at } 26\% \text{ QCD eff.})$$

The **scaling variable** for QCD jets ρ is used in the characterization of the correlation of jet substructure variable

Jet Substructure

ArXiv:1603.00027
CMS-PAS-B2G-17-001
CMS-PAS-EXO-17-001

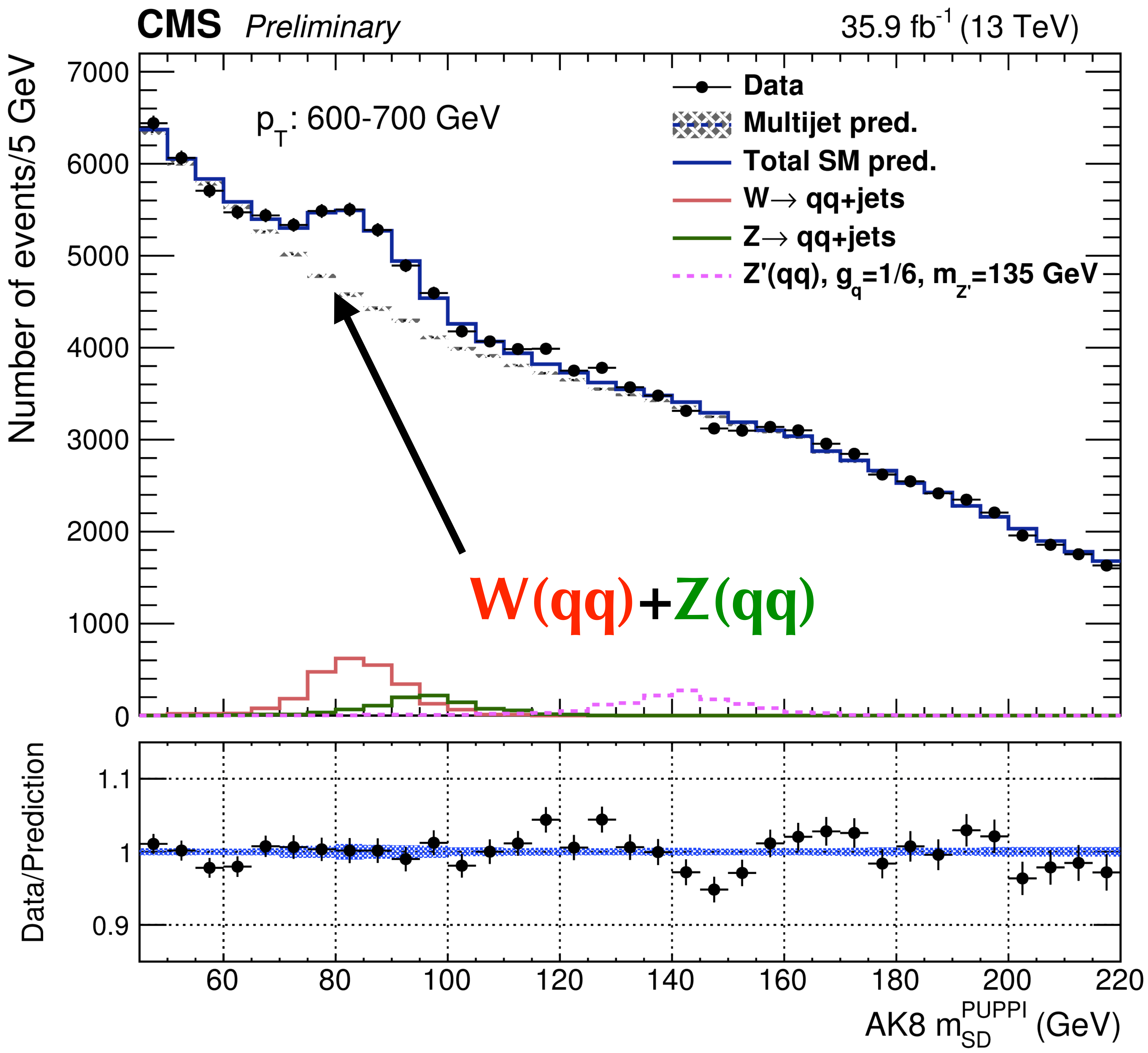


N_2 sculpts the jet mass distribution

We use a **mass-decorrelated** version for the background

Jet Substructure

ArXiv:1603.00027
CMS-PAS-B2G-17-001
CMS-PAS-EXO-17-001



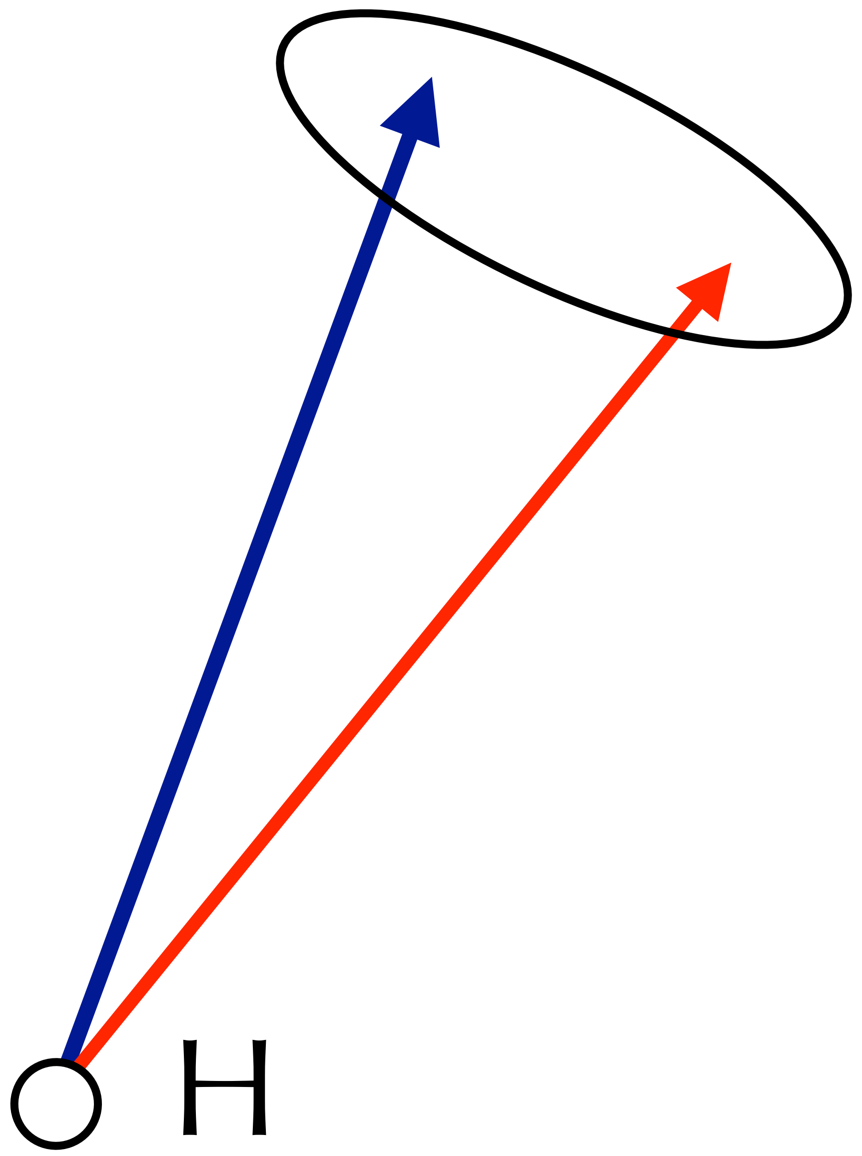
N_2 sculpts the jet mass distribution

We use a **mass-decorrelated** version for the background

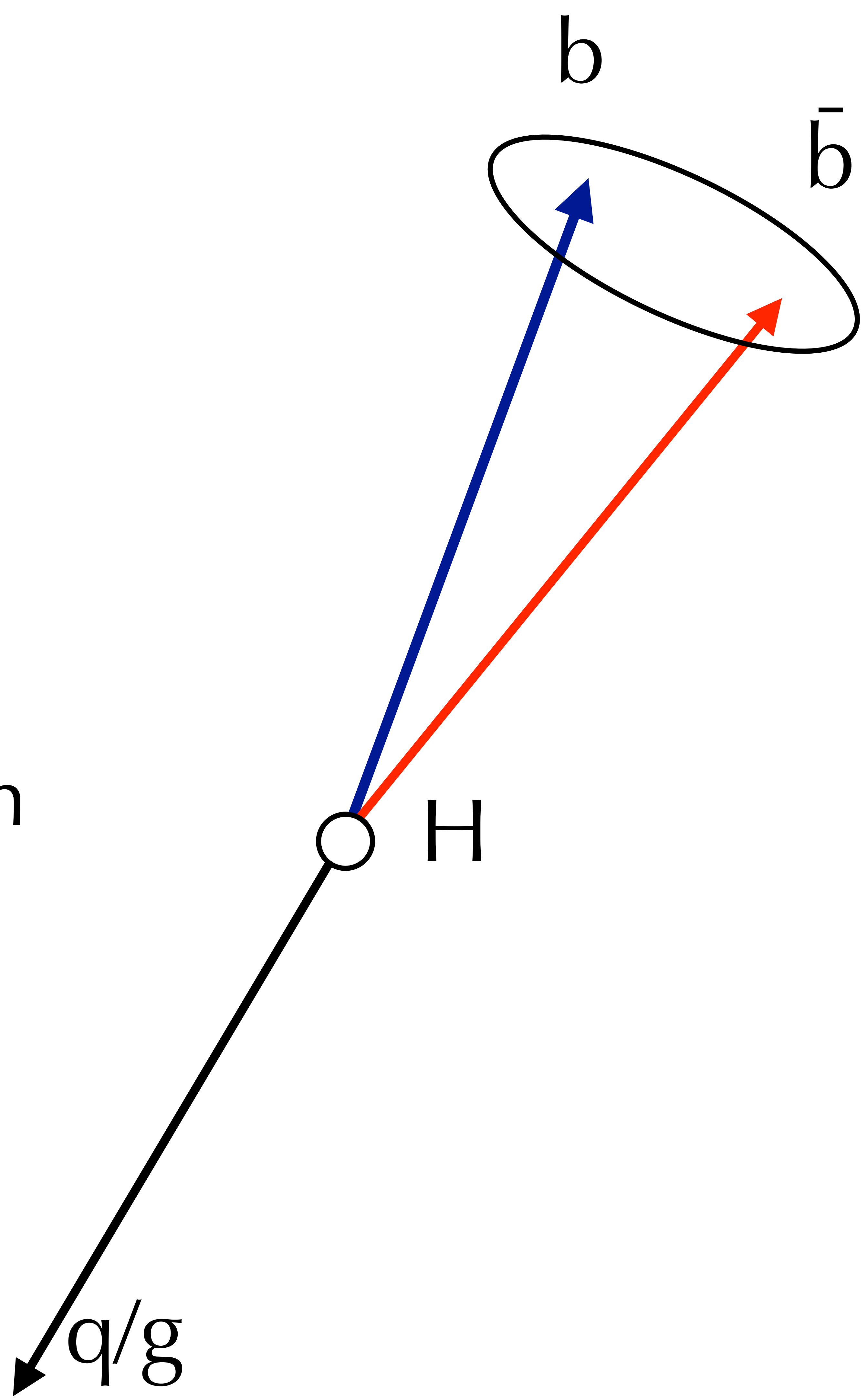
Phys. Rev. Lett. 100 (2008) 242001

> *We conclude that subjet techniques have the potential to transform the high-p_T WH, ZH($H \to b\bar{b}$) channel into one of the best channels for discovery of a low mass Standard Model Higgs at the LHC.*

now we search for <u>inclusive</u> H to bb with these new tools ...



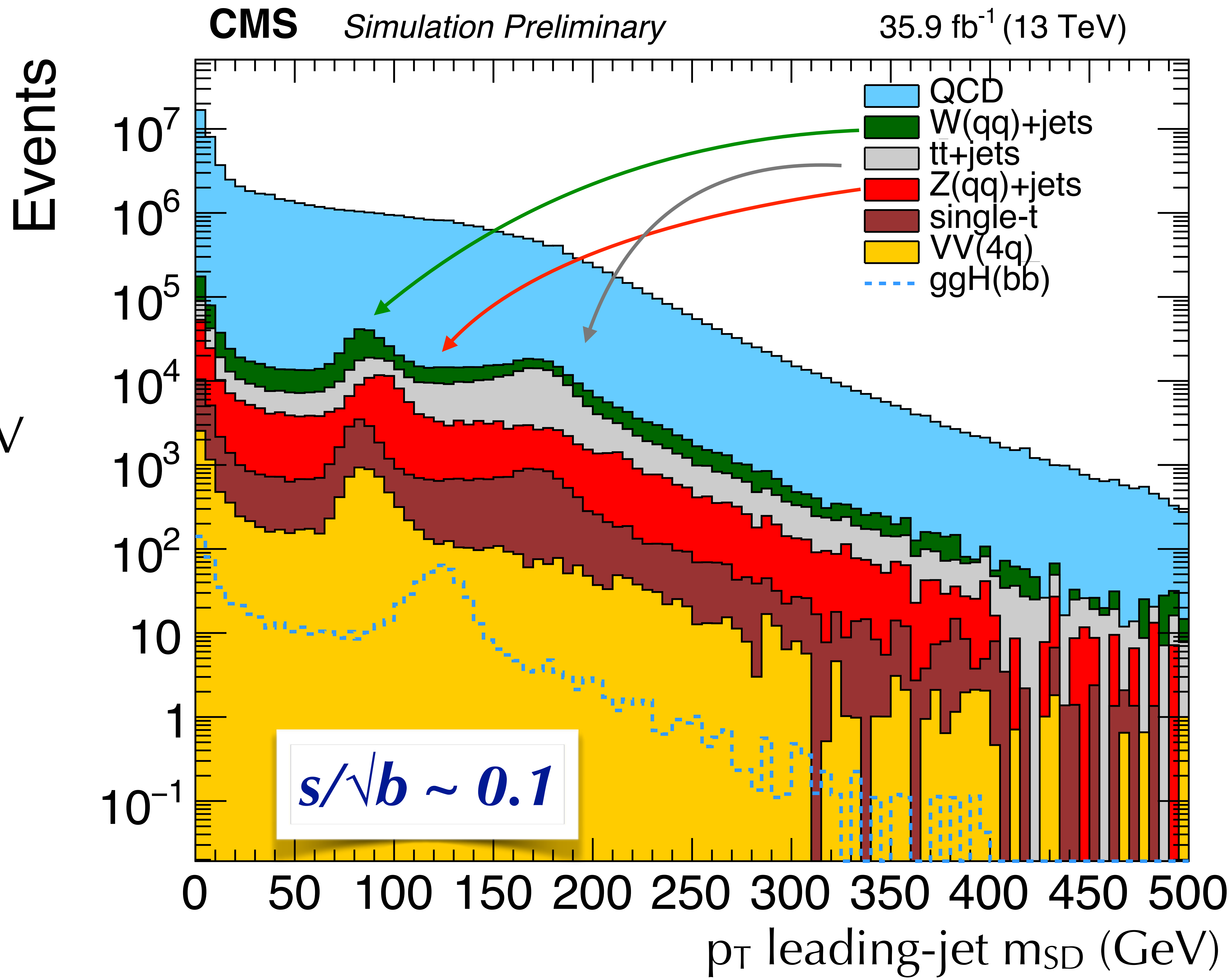
Search for inclusive H to $b\bar{b}$



- We can access this process in the **boosted dijet** topology
- Use initial state jet to get above the trigger threshold
- Look for boosted **H boson in a single jet mass** distribution
  - Use the **Z boson as Standard Model** candle
    - b-tagging to disentangle W/Z

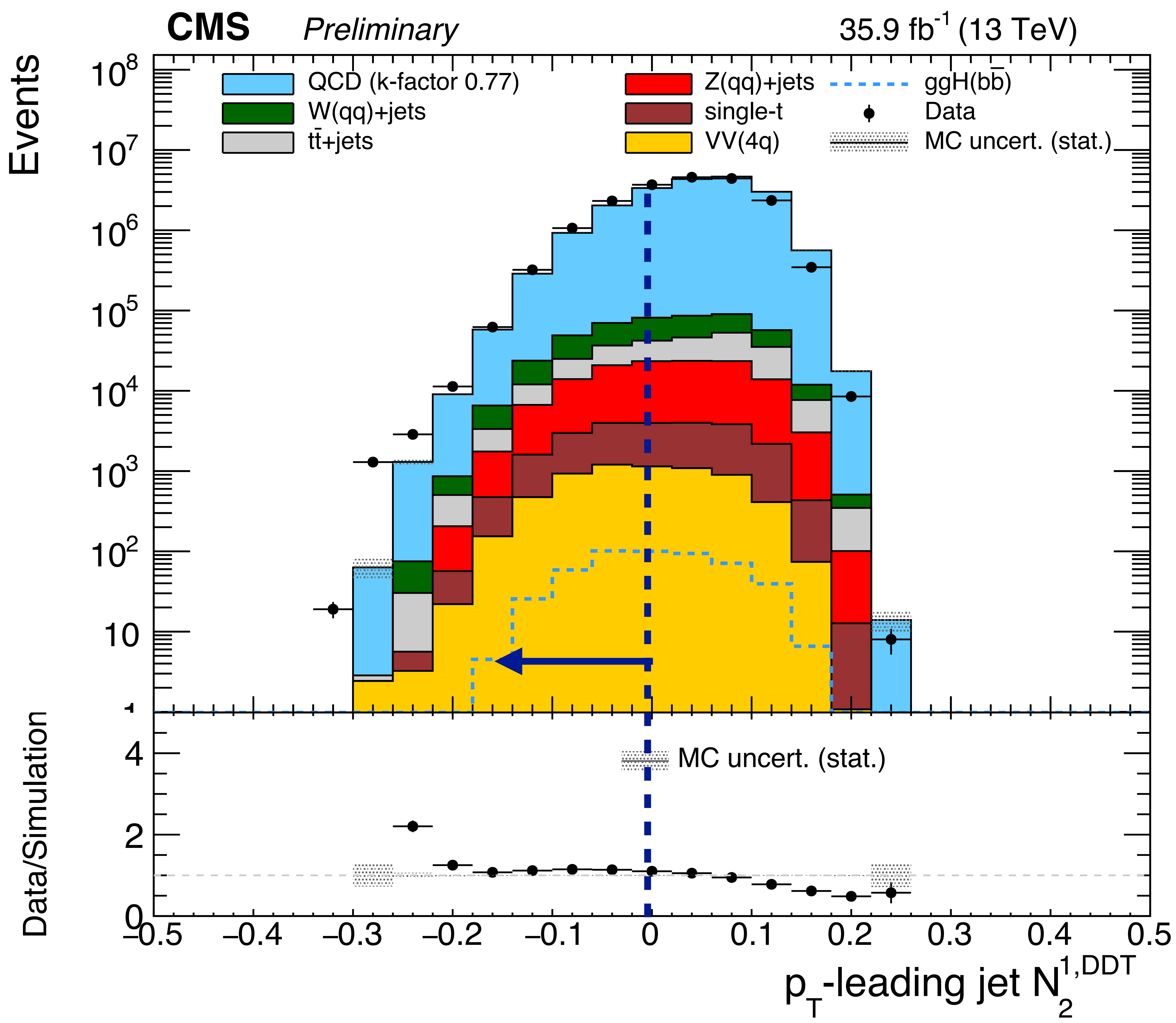
Event Selection

- **Online selection** asks for a high p_T single jet or large hadronic activities
  - $p_T > 360$ GeV or $\Sigma\, p_T > 800/900$ GeV
- Offline: **Highest p_T jet**:
  - $p_T > 450$ GeV $|\eta| < 2.5$,
  - jet soft drop mass (**m_{SD}**) > 40 GeV
- Lepton veto, E_T^{Miss} veto

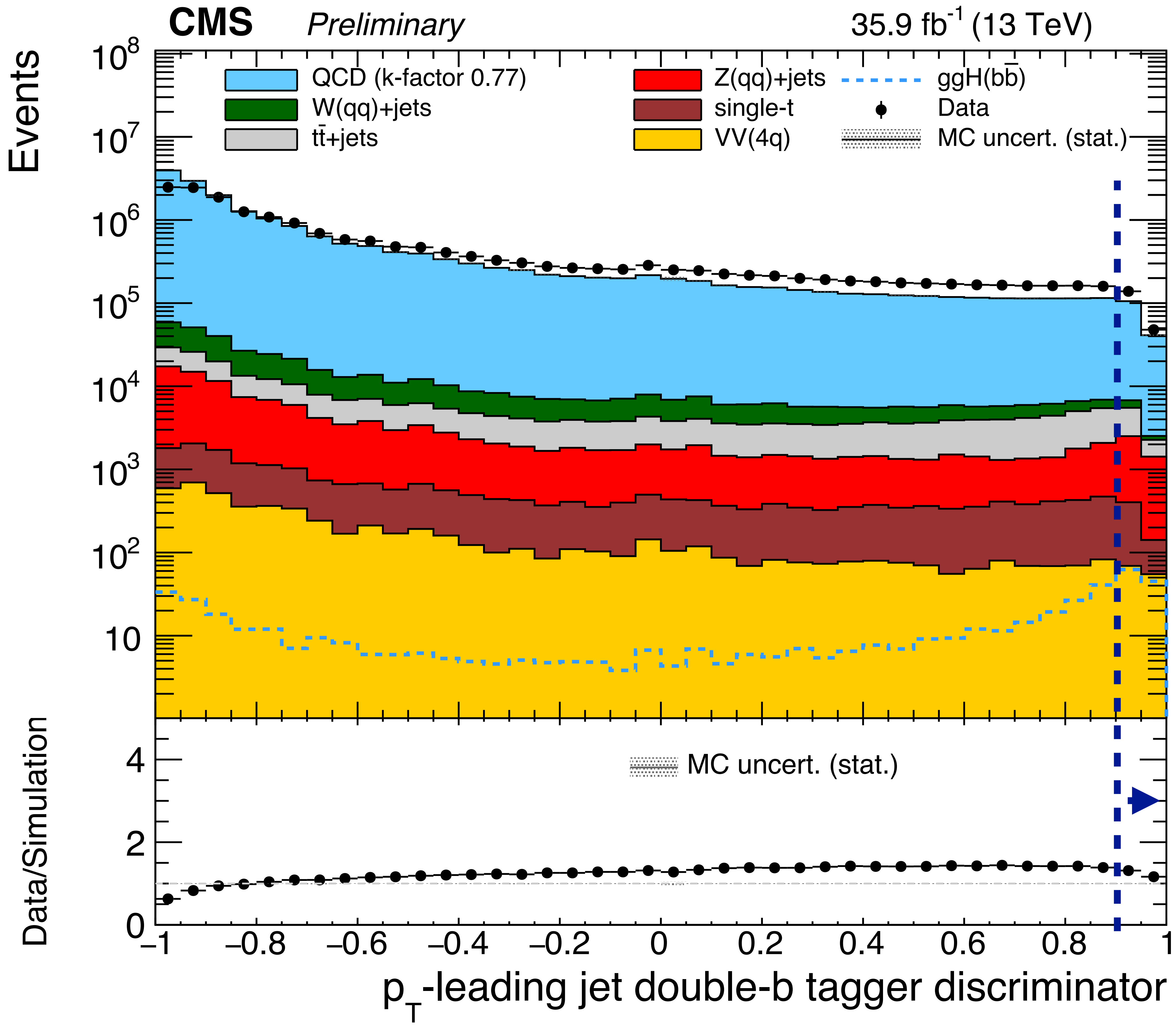


Event Selection

Substructure: two prongs discrimination

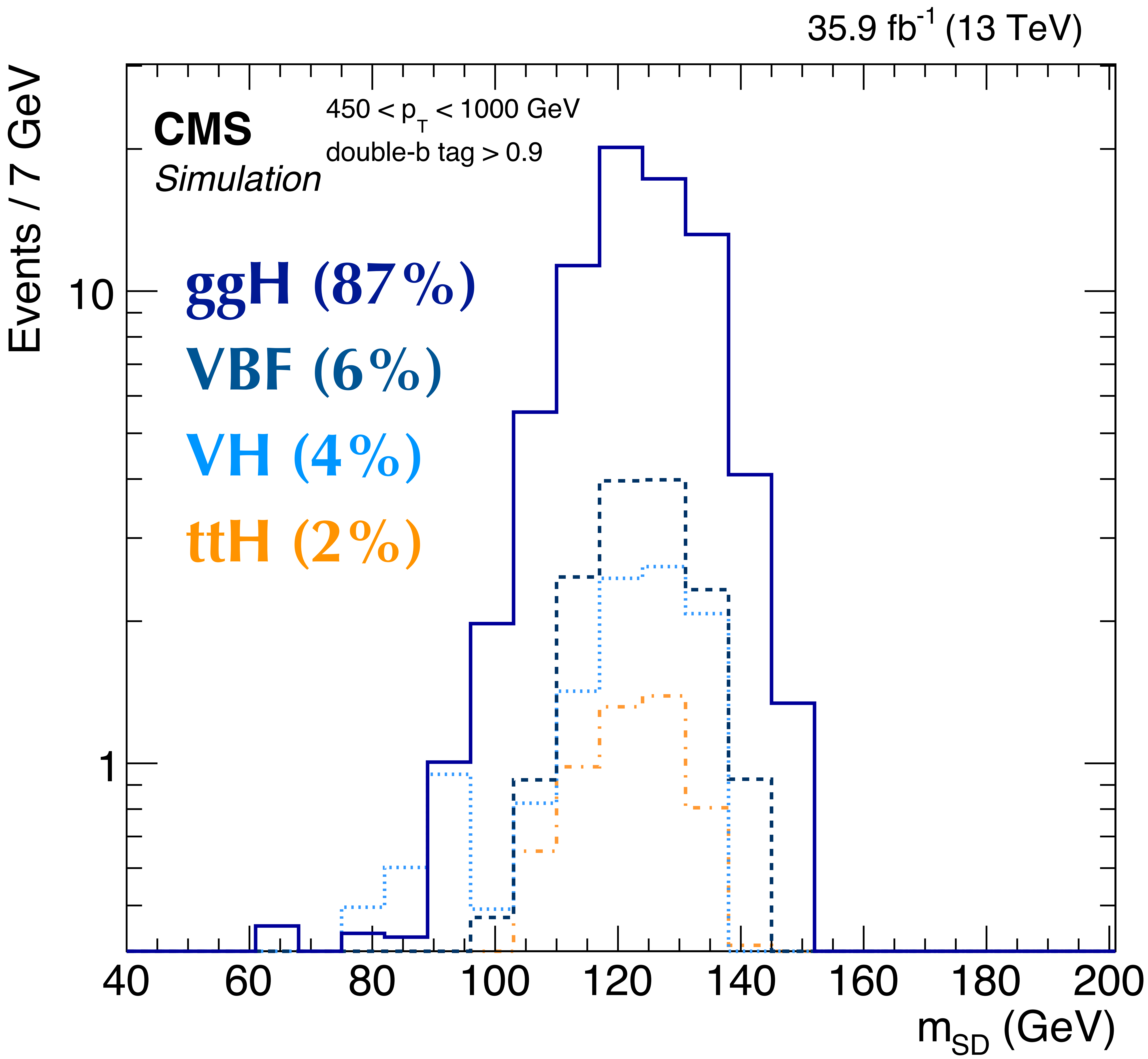


double-b tagger

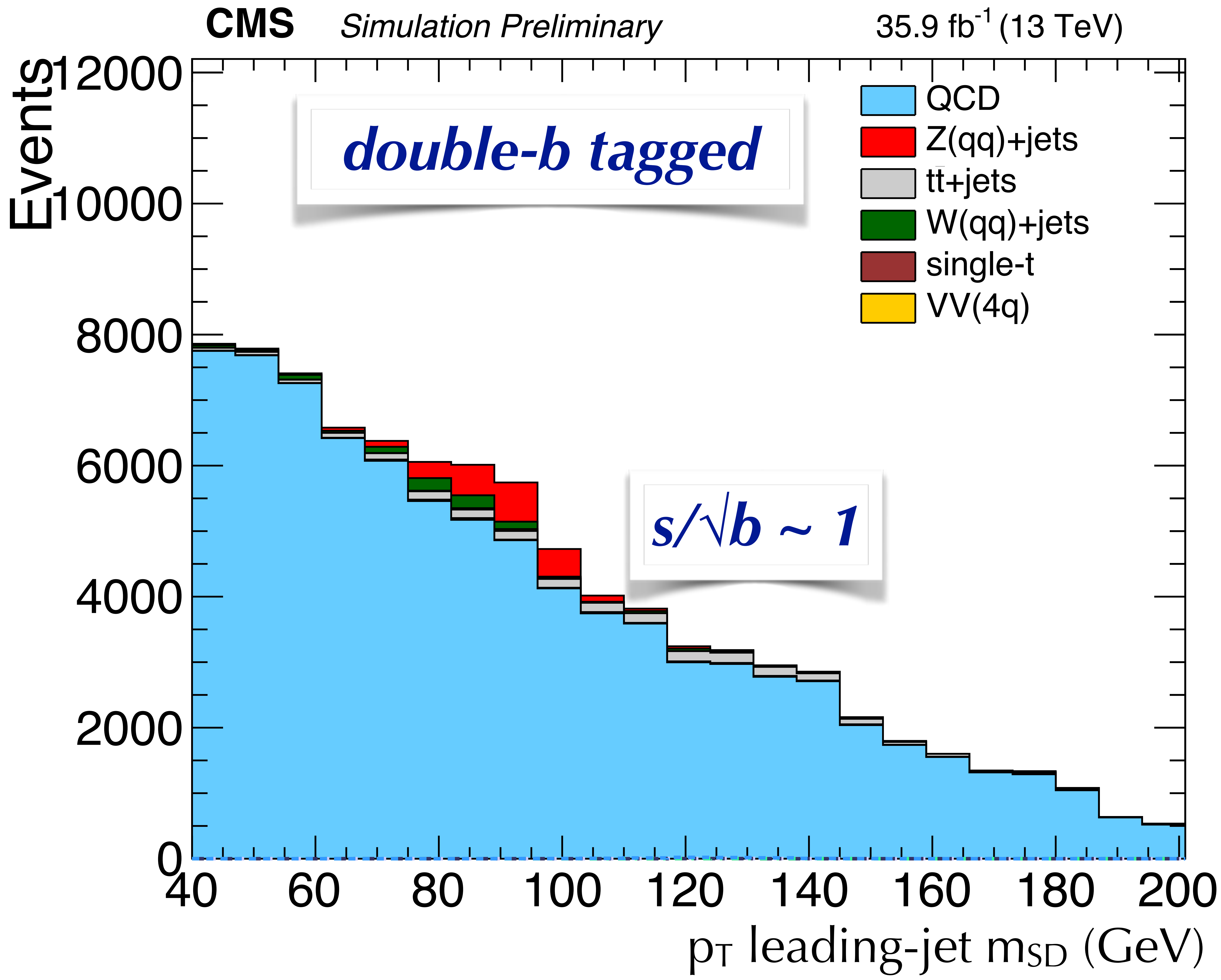


After all selections

- Analysis is inclusive in Higgs production mode
- Dominant contribution in signal region is ggH



Background composition



Backgrounds

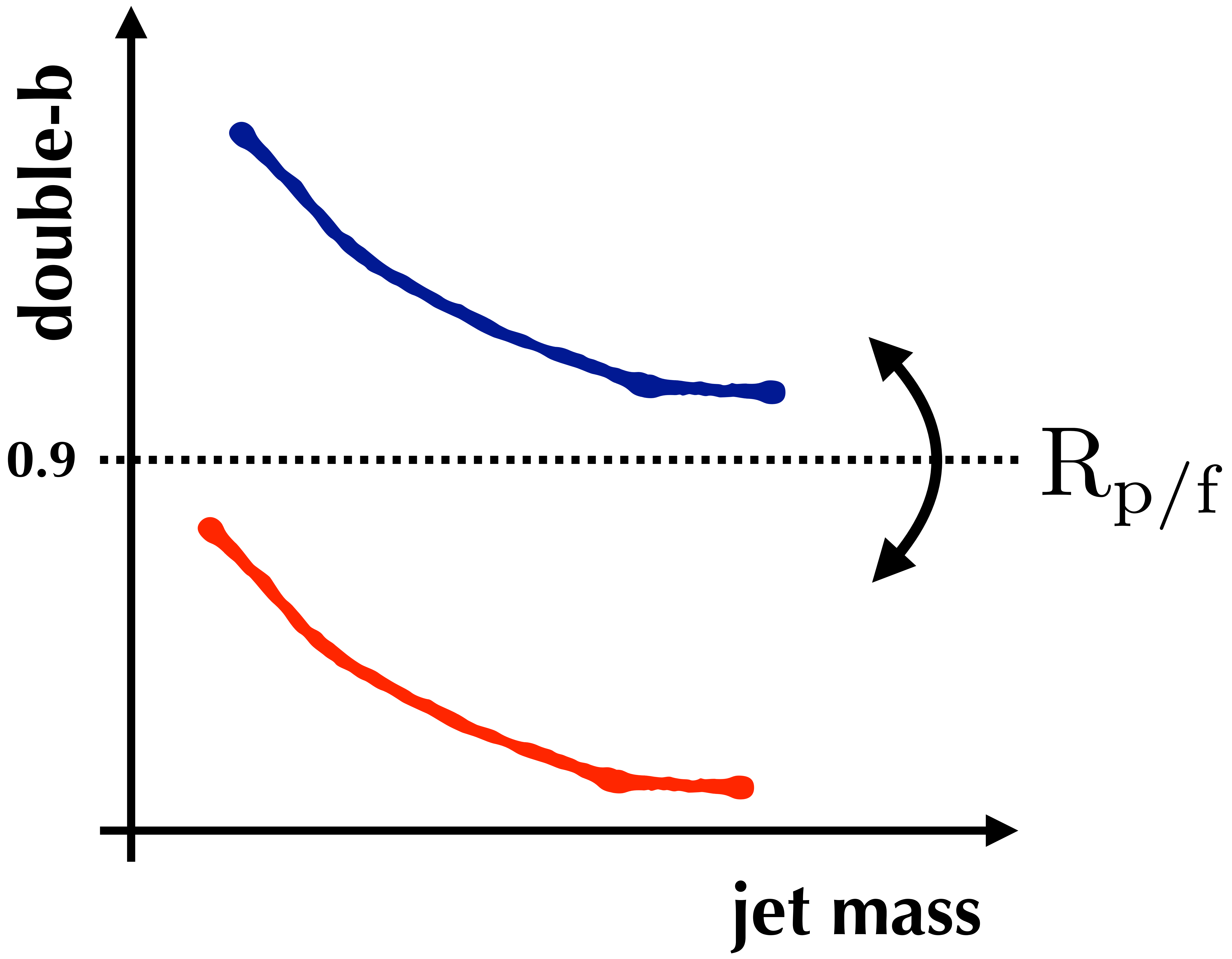
- QCD (~**90%**)
- tt+jets (**3%**), normalization from a dedicated control region

} **data**

- W/Z+jets (**5%**)
- single-t, VV (<**1%**)

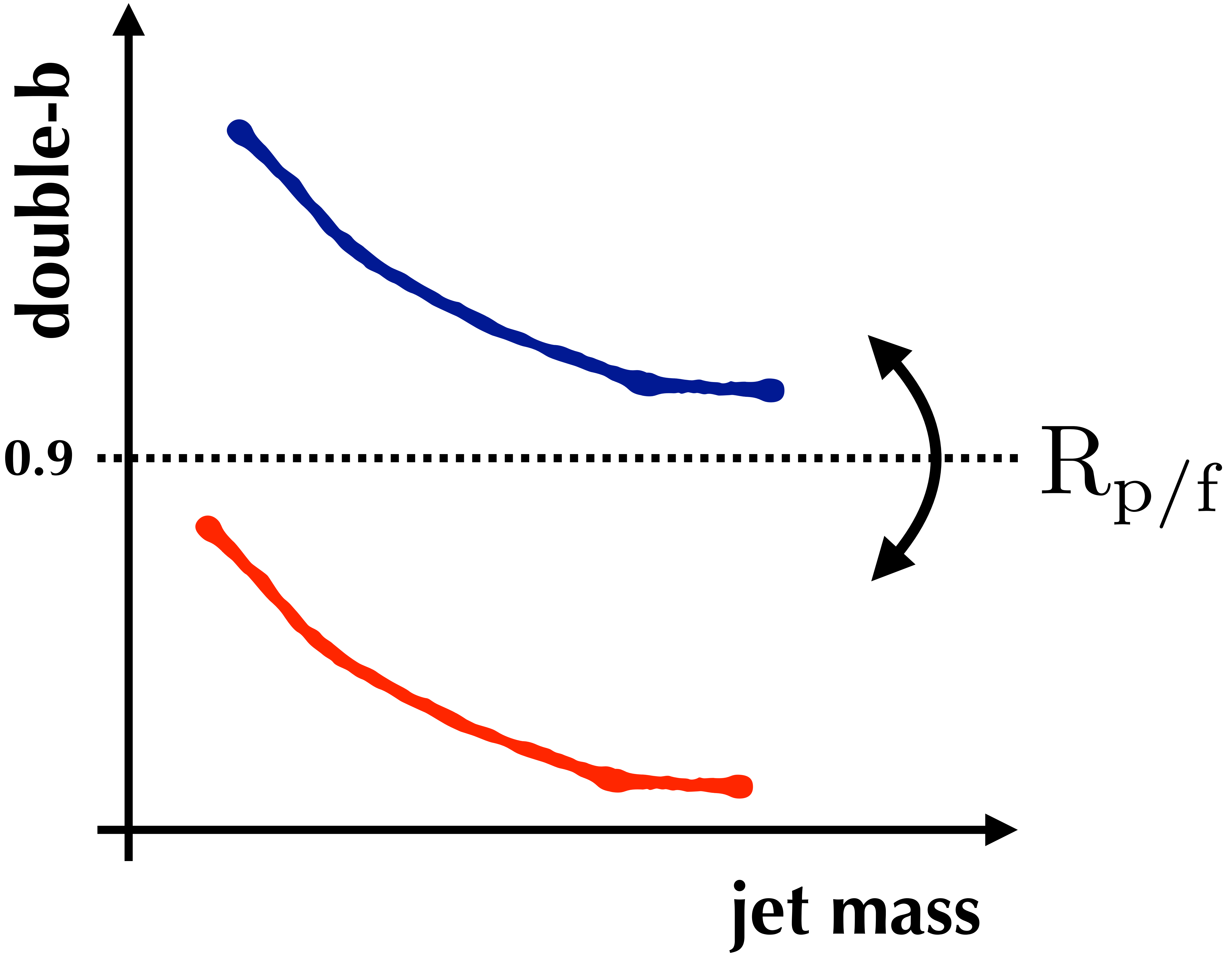
} **simulation**

Multijets background prediction

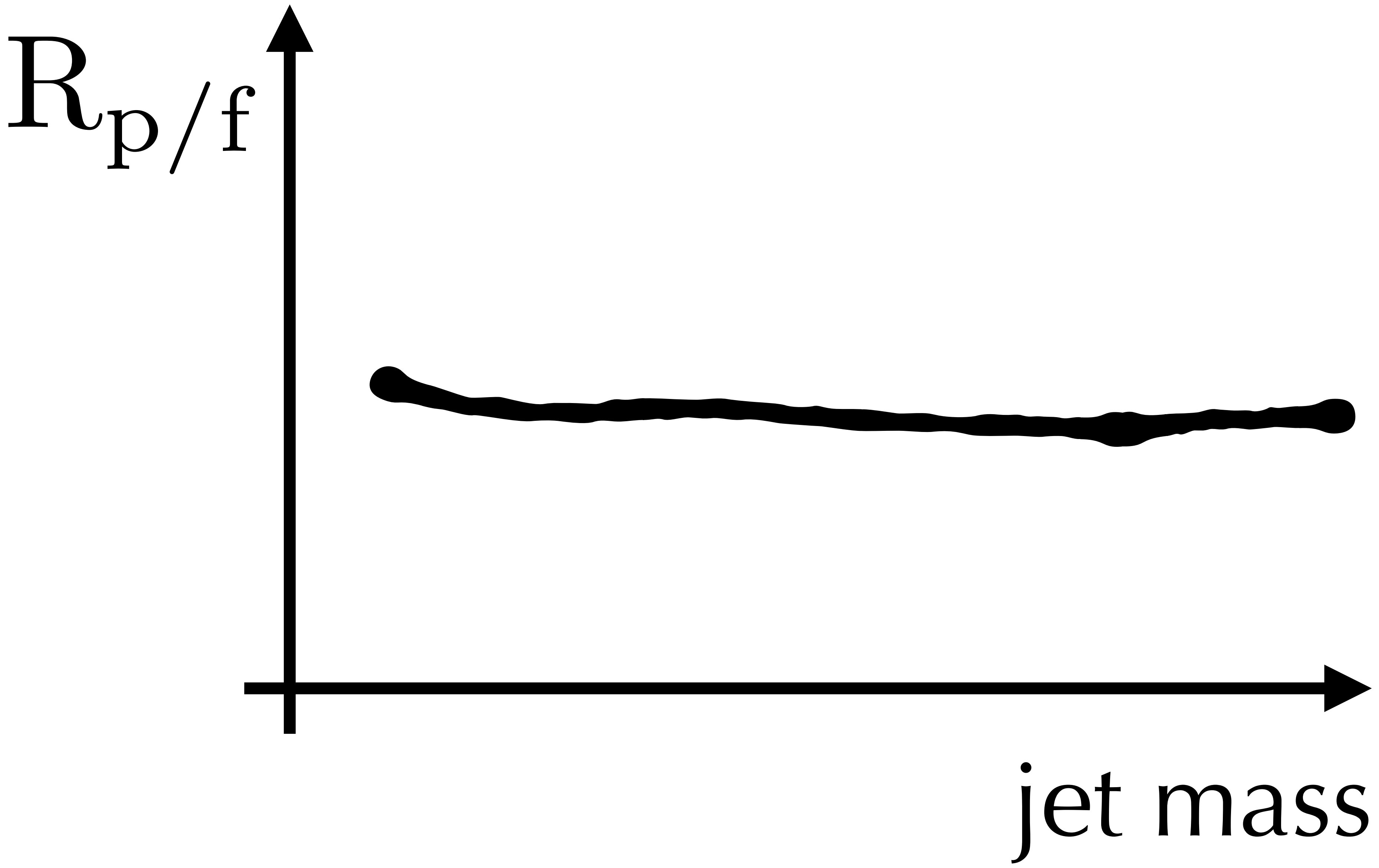


- Jet mass shape for multijets events is derived in data
- From events “failing” the b-tag requirement with a ***transfer factor***

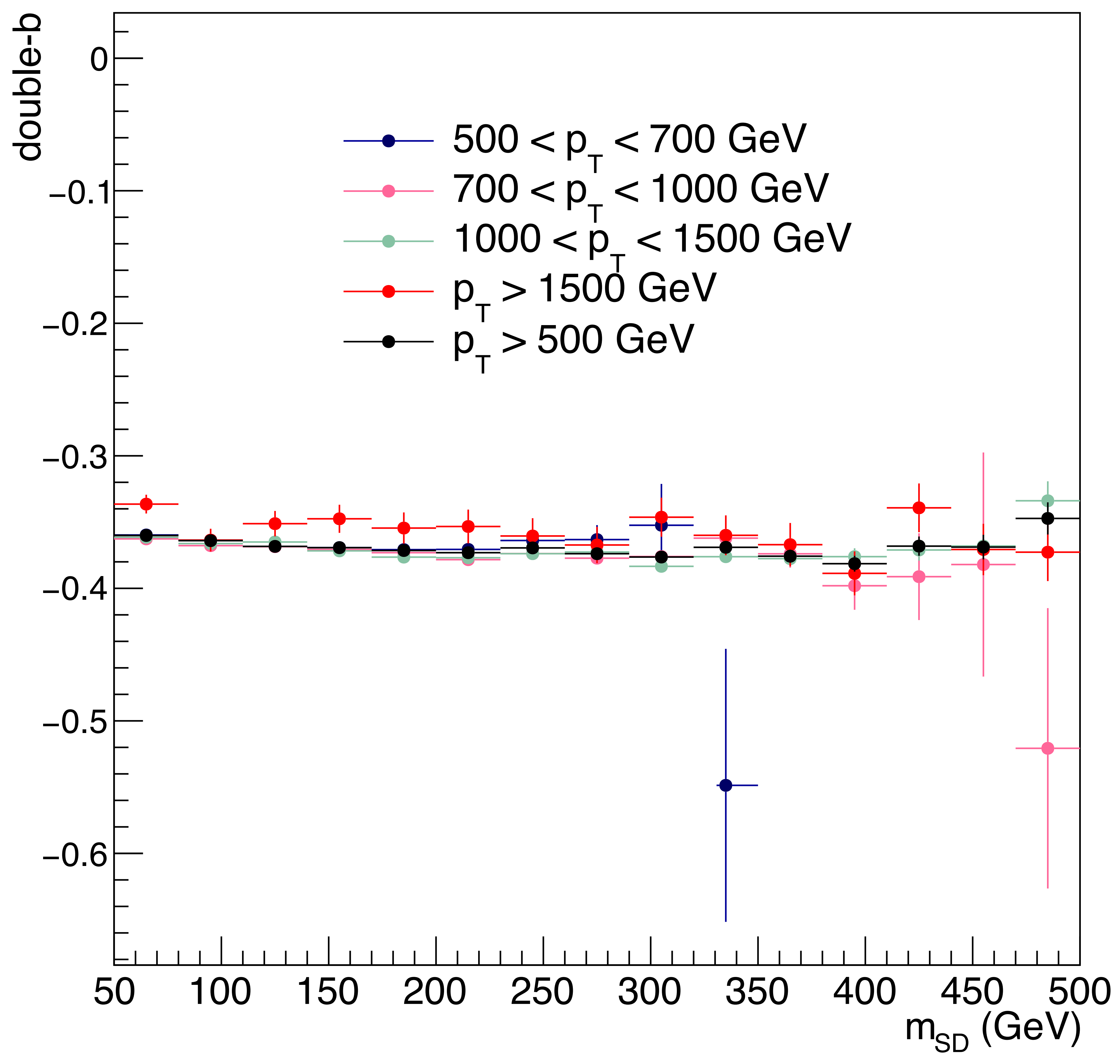
Multijets background prediction



- Jet mass shape for multijets events is derived in data
- From events “failing” the b-tag requirement with a ***transfer factor***



Multijets background prediction

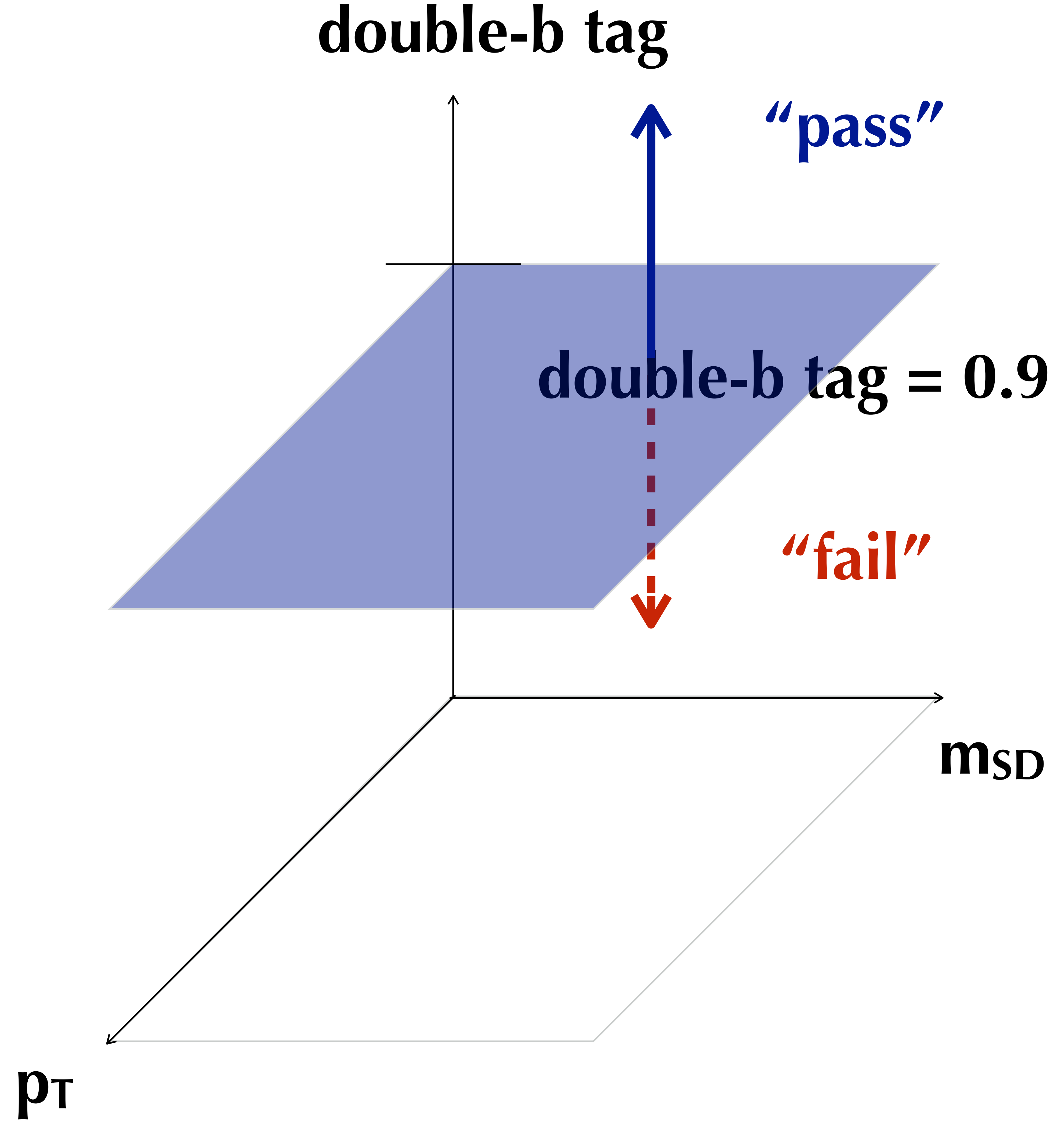


- Jet mass shape for multijets events is derived in data
- From events "failing" the b-tag requirement with a ***transfer factor***
  - as function of the **jet mass and p_T**

Multijets background prediction

- Jet mass shape for multijets events is derived in data
- From events "failing" the b-tag requirement with a ***transfer factor*** as function of the **jet mass and p_T**

$$N^{QCD}_{pass}(m_{SD}, p_T) = R_{p/f}(\rho, p_T) \times N^{QCD}_{fail}(m_{SD}, p_T)$$

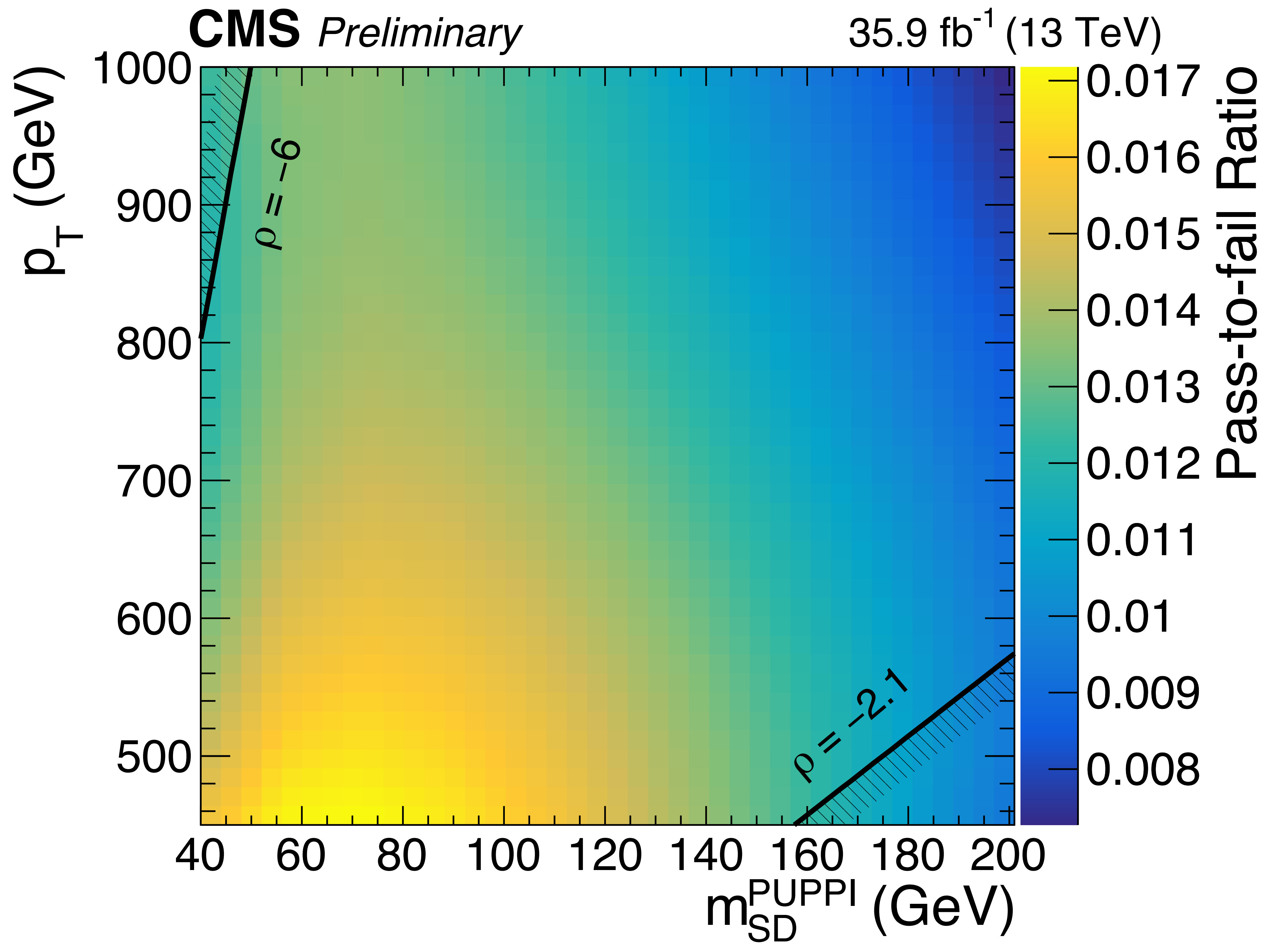


Multijets background prediction

- Jet mass shape for multijets events is derived in data
- From events "failing" the b-tag requirement with a ***transfer factor*** as function of the **jet mass and $\mathbf{p_T}$**

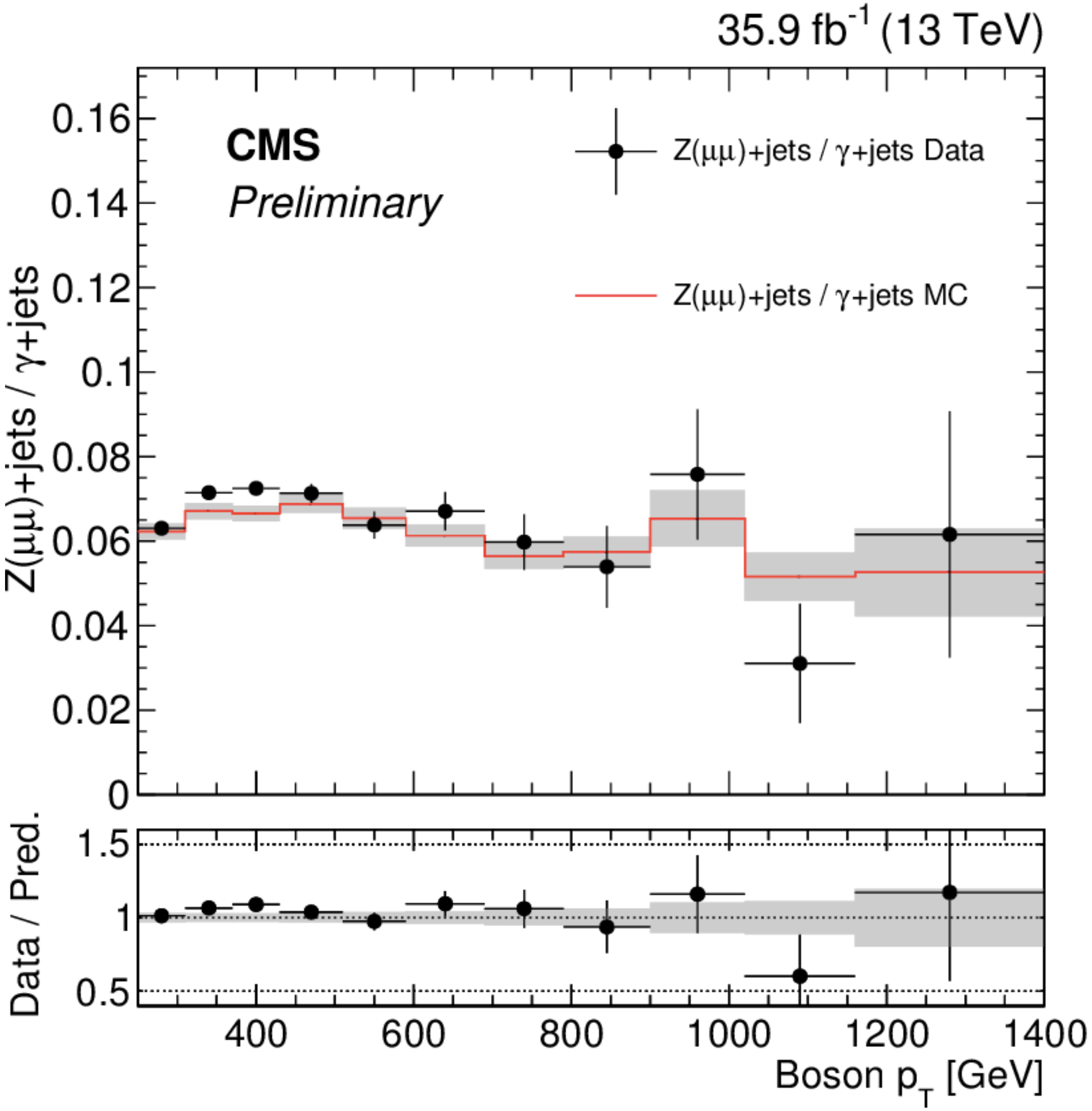
$$N^{QCD}_{pass}(m_{SD}, p_T) = R_{p/f}(\rho, p_T) \times N^{QCD}_{fail}(m_{SD}, p_T)$$

- The transfer factor is determined **simultaneously with the signal extraction**



W/Z+jets simulation

ArXiv:1705.04664
CMS-PAS-EXO-16-048

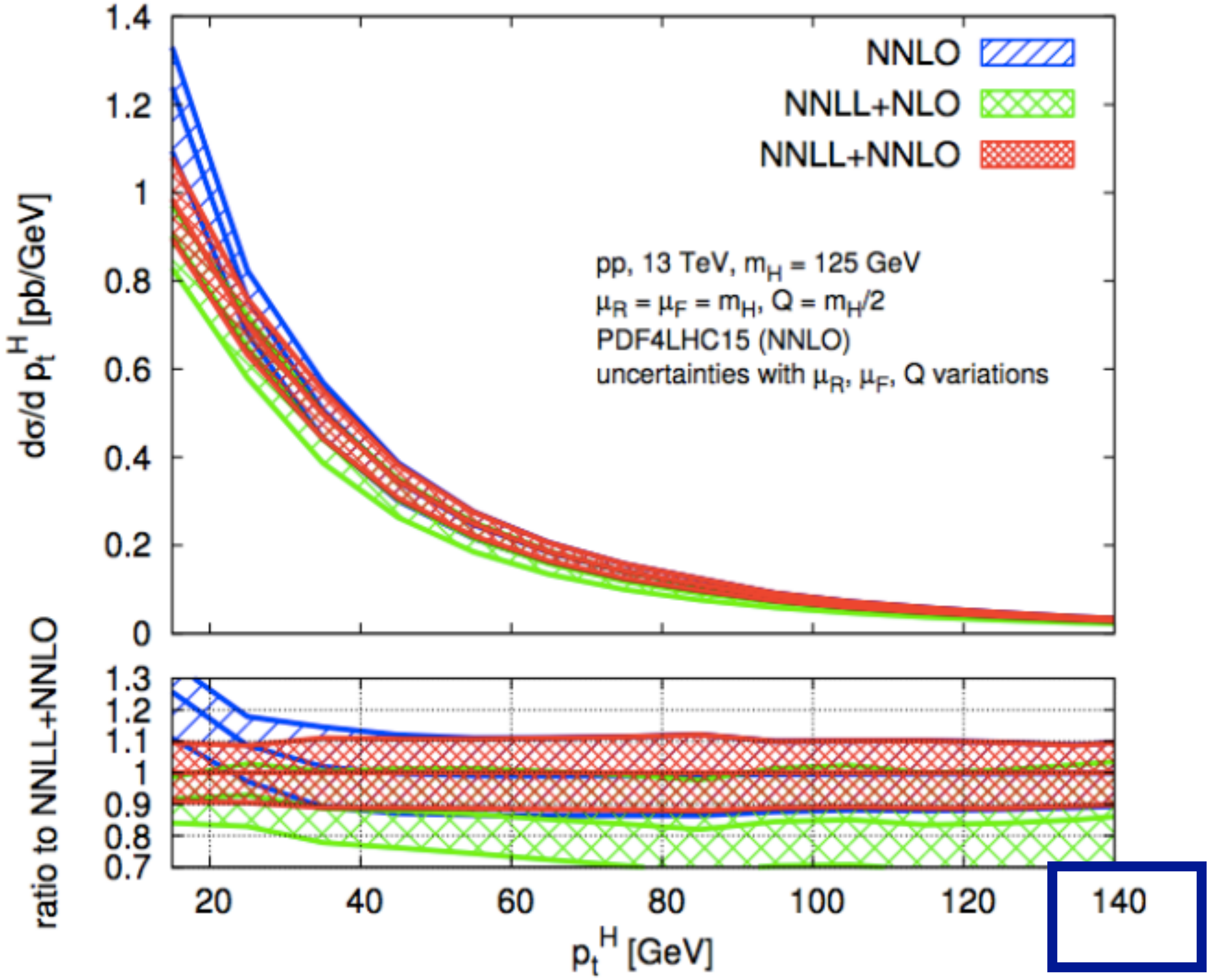


- LO simulations for the W/Z+jets are corrected using **p_T-dependent** :
  - **NLO QCD** k-factors
  - **NLO electroweak** k-factors
- Associated uncertainties are 10% (QCD) and 15-35% (EWK)

ArXiv:1410.5806, ArXiv:1609.00367
ArXiv:1408.5325, ArXiv:1302.6216
ArXiv:1504.07922, ArXiv:1505.03893
ArXiv:1610.07922

ggH simulation at high p_T

- Other CMS Higgs results use **Powheg** : 1 jet + $m_t = \infty$
- We want to account for both the effects of **higher order corrections** and for the **finite top mass**



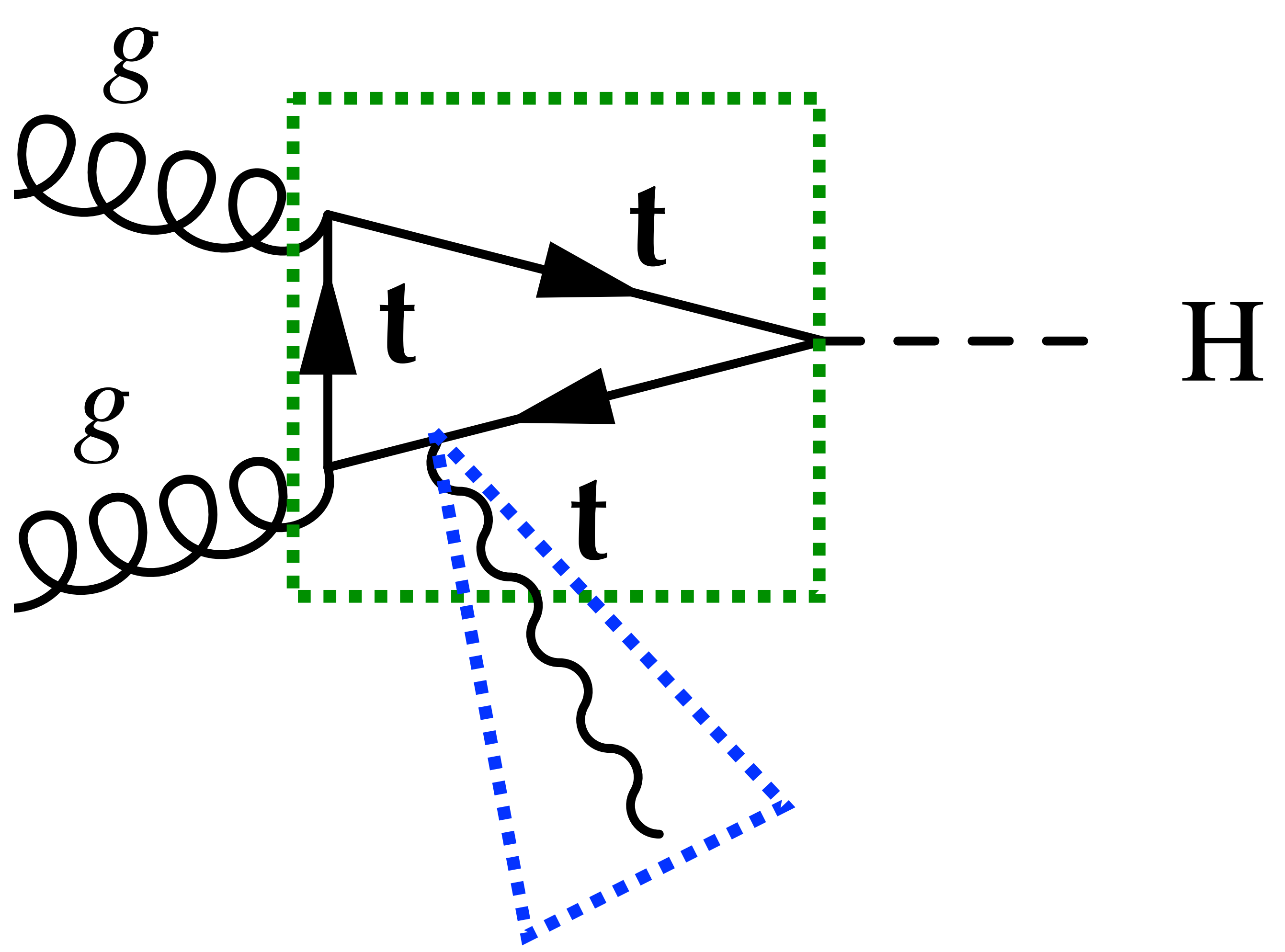
≪1000 GeV

No real NLO + finite m_t calculation available in the literature above p_T^H > 300 GeV

ggH simulation at high p_T

ArXiv:1410.5806, ArXiv:1609.00367
ArXiv:1408.5325, ArXiv:1302.6216
ArXiv:1504.07922, ArXiv:1505.03893
ArXiv:1610.07922

- Other CMS Higgs results use **Powheg** : 1 jet + $m_t = \infty$
- We want to account for both the effects of **higher order corrections** and for the **finite top mass**



A multi-correction approach is adopted

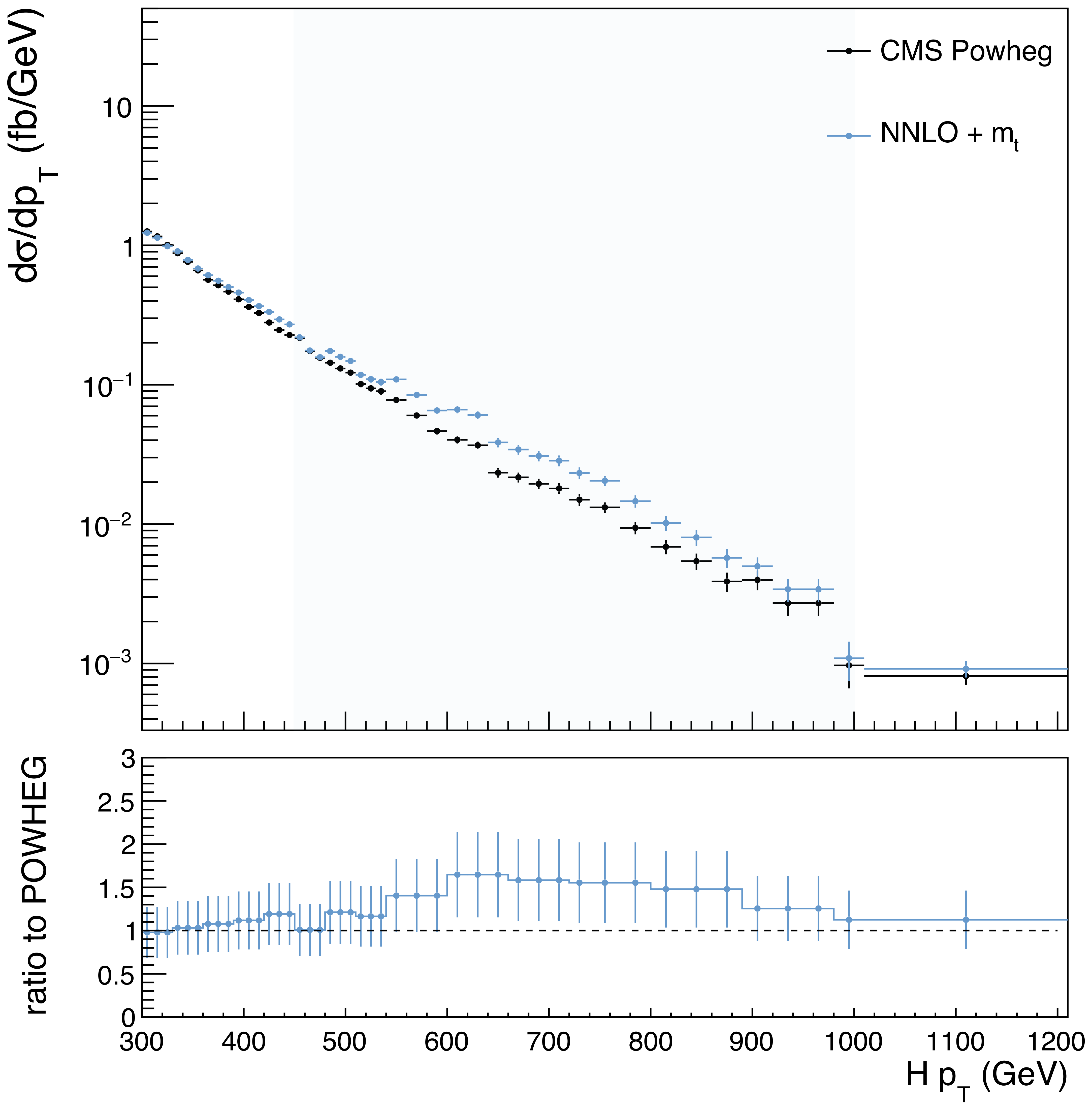
- LO H+0-2jet, finite m_t
- NLO H+1jet finite m_t up to $1/m_t^4$ expansion
- NNLO H+1jet, $m_t = \infty$

$$\mathrm{GF\ H(NNLO + m_t) = (1\ jet\ m_t = \infty) \times \frac{MG\ LO\ 0-2\ jet\ m_t}{(1\ jet\ m_t = \infty)} \times \frac{NLO\ 1\ jet\ m_t}{LO\ 1\ jet\ m_t} \times \frac{NNLO\ 1\ jet\ m_t = \infty}{NLO\ 1\ jet\ m_t = \infty}}$$

ggH p_T reweighting finite m_t + NNLO

- This is the **first time an (approximate) NLO H+0,1,2 jet merged with finite top mass** is attempted
- Estimate k-factor of ~1.3 for H p_T > 450 GeV

Results are stable under these variations and also provided without p_T reweighting



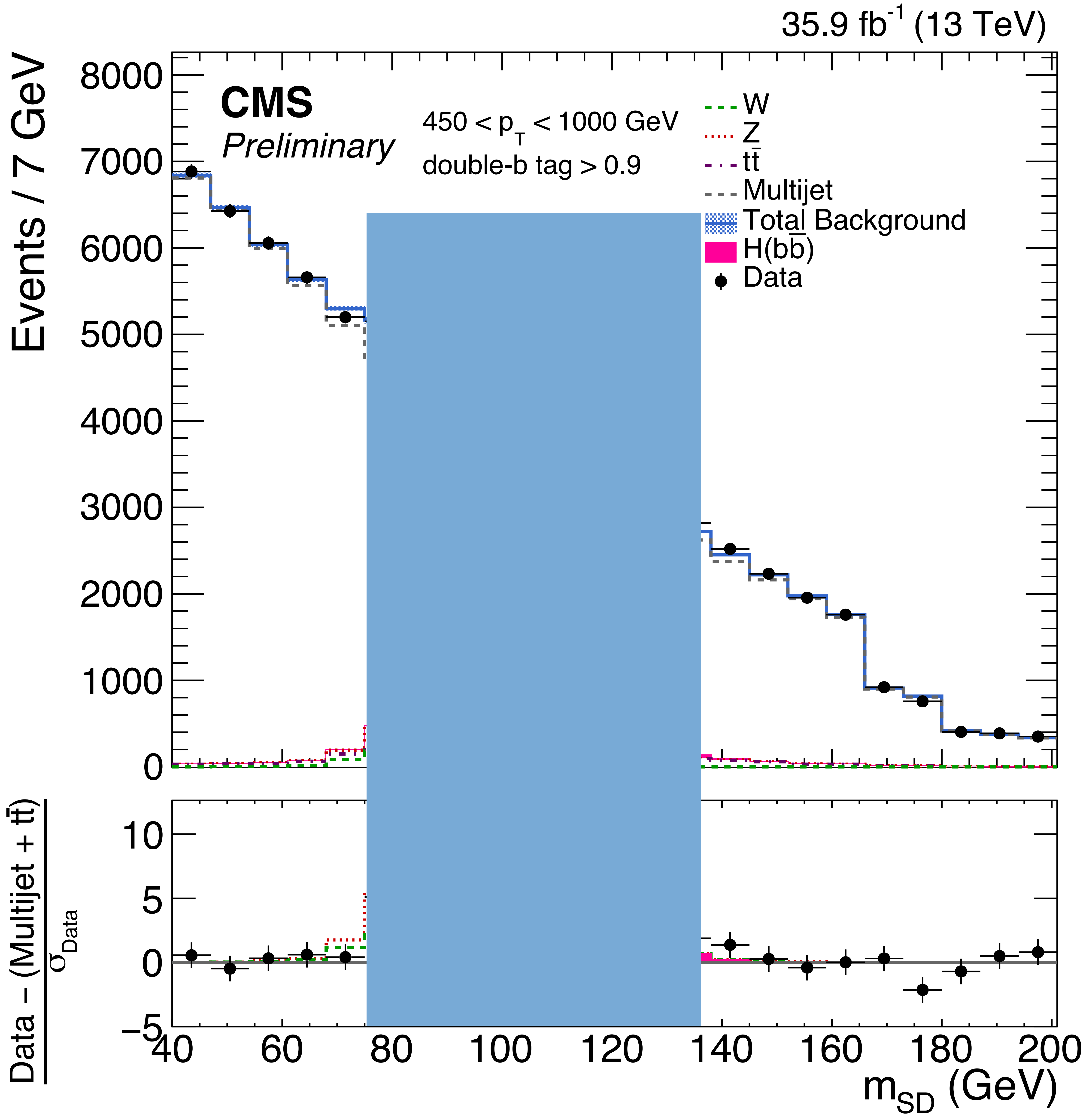
Systematics

Systematic uncertainty source	Type (shape or normalization)	Relative size (or description)
QCD transfer factor	both	profile $a_{k\ell}$ and QCD normalization
Luminosity	normalization	2.5%
V-tag ($N_2^{1,DDT}$) efficiency	normalization	4.3%
Muon veto efficiency	normalization	0.5%
Electron veto efficiency	normalization	0.5%
Trigger efficiency	normalization	4%
Muon ID efficiency	shape	up to 0.2%
Muon isolation efficiency	shape	up to 0.1%
Muon trigger efficiency	shape	up to 8%
$t\bar{t}$ normalization SF	normalization	from 1μ CR: 8%
$t\bar{t}$ double-b mis-tag SF	normalization	from 1μ CR: 15%
W/Z NLO QCD corrections	normalization	10%
W/Z NLO EWK corrections	normalization	15% – 35%
W/Z NLO EWK ratio decorrelation	normalization	5% – 15%
double-b tagging efficiency	normalization	4%
Jet energy scale	normalization	up to 10%
Jet energy resolution	normalization	up to 15%
Jet mass scale	shape	shift m_{SD} peak by $\pm 0.4\%$
Jet mass resolution	shape	smear m_{SD} distribution by $\pm 9\%$
Jet mass scale p_T	normalization	0.4%/100 GeV (p_T)
Monte Carlo statistics	normalization	-
H p_T correction (gluon fusion)	both	30%

correlated among W, Z, and H

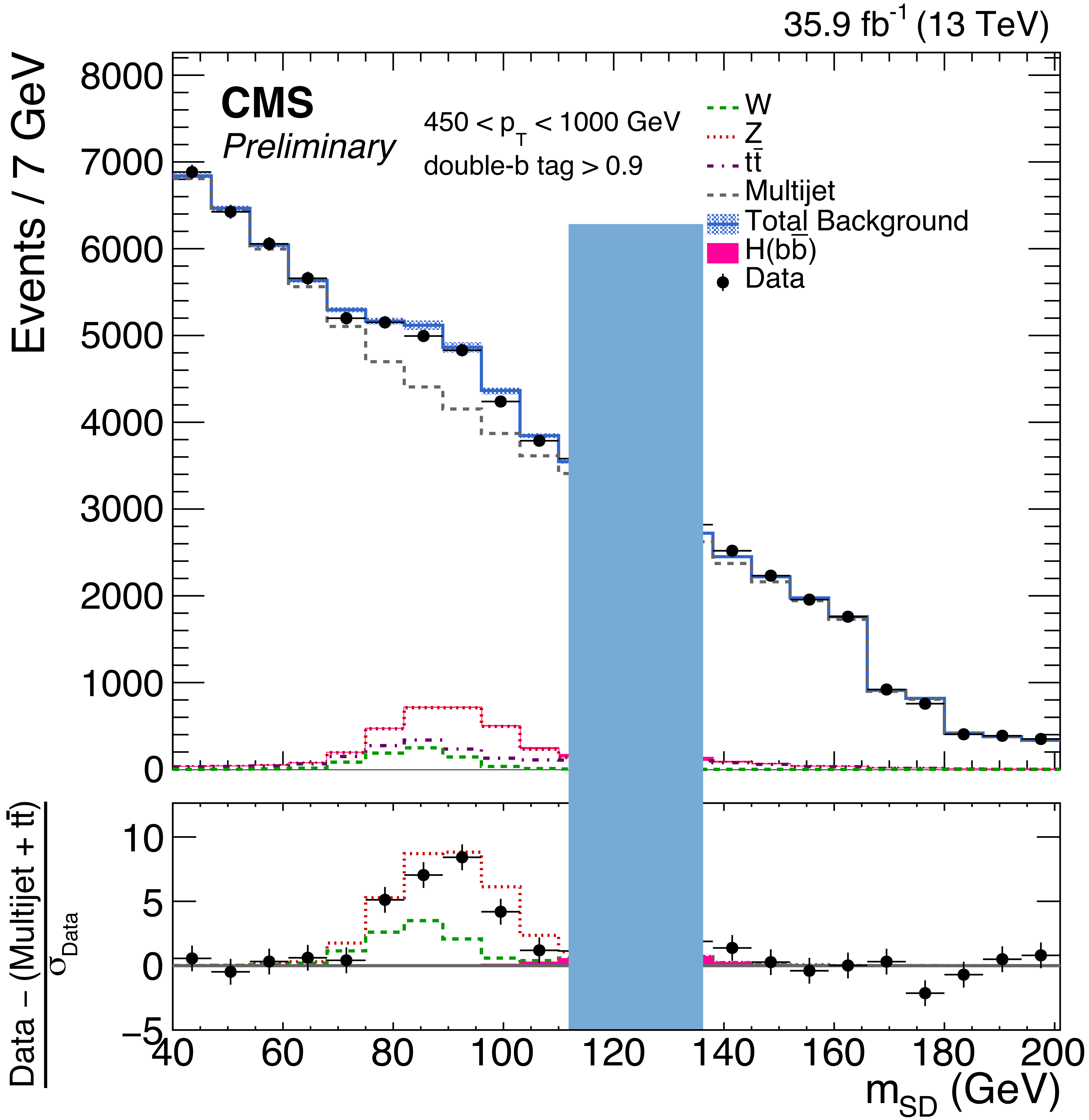
Z signal extraction

- The extraction of the Z signal compatible with the SM expectation would validate the H signal extraction and H-tagging approach

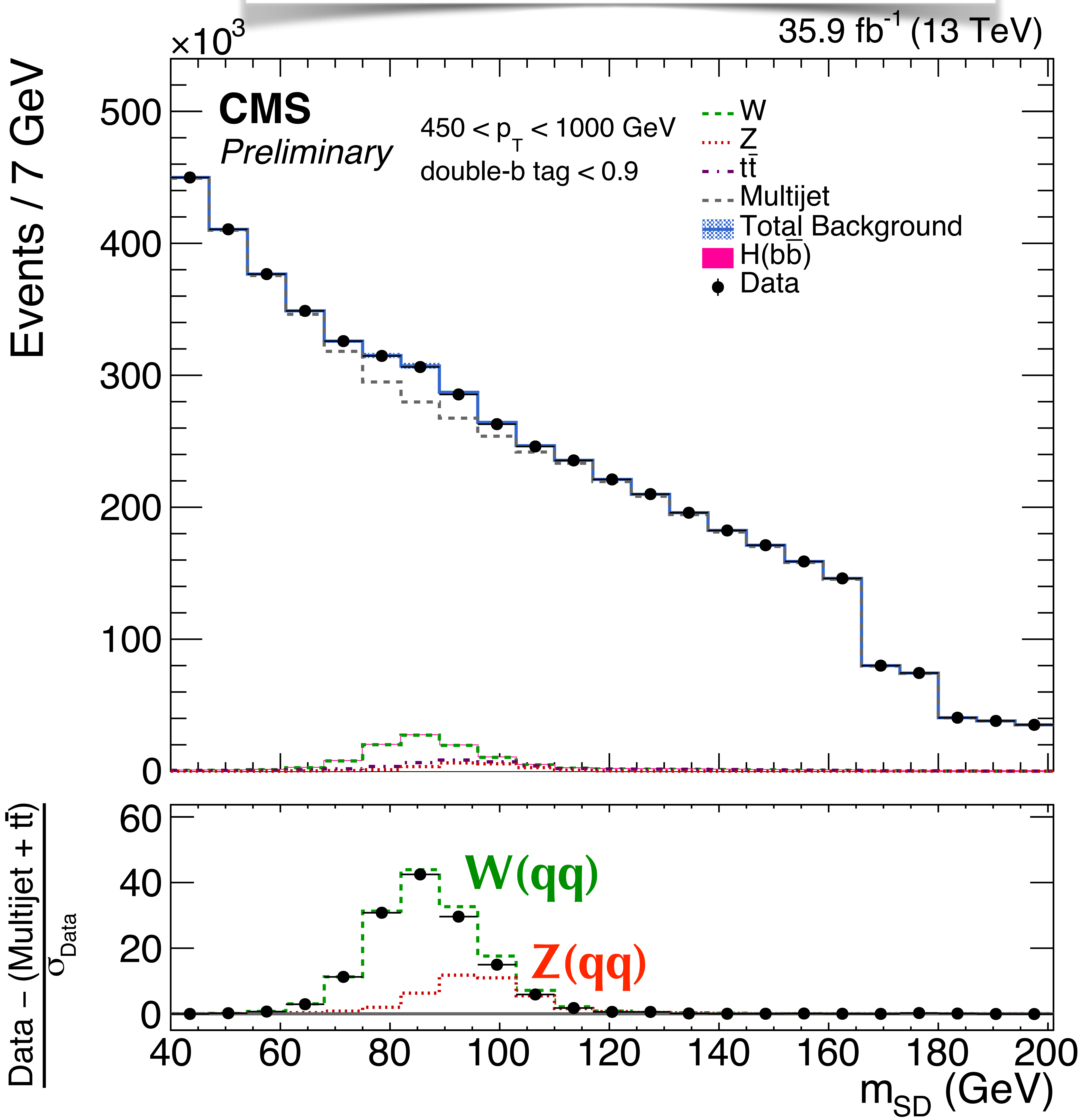


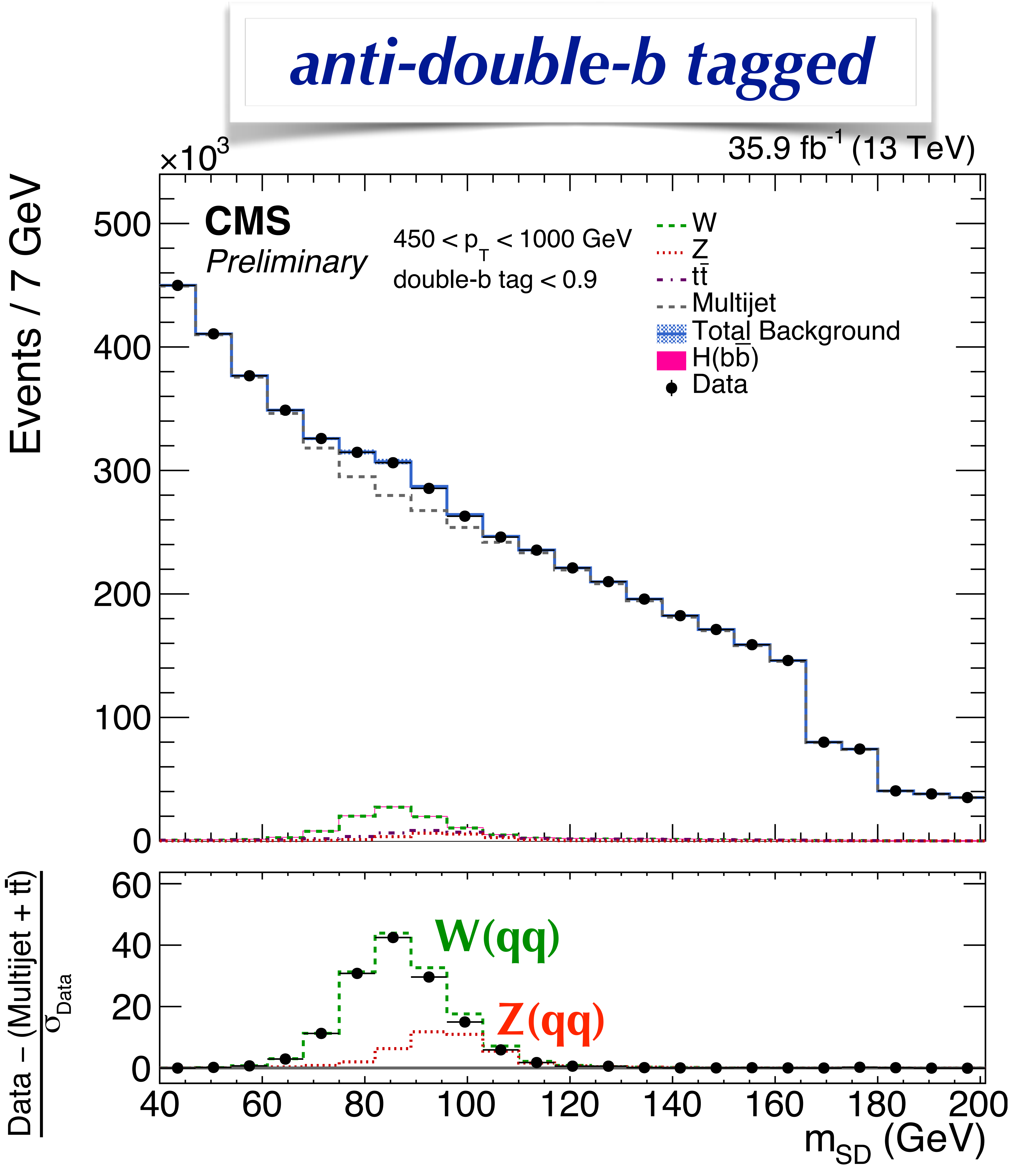
Z results

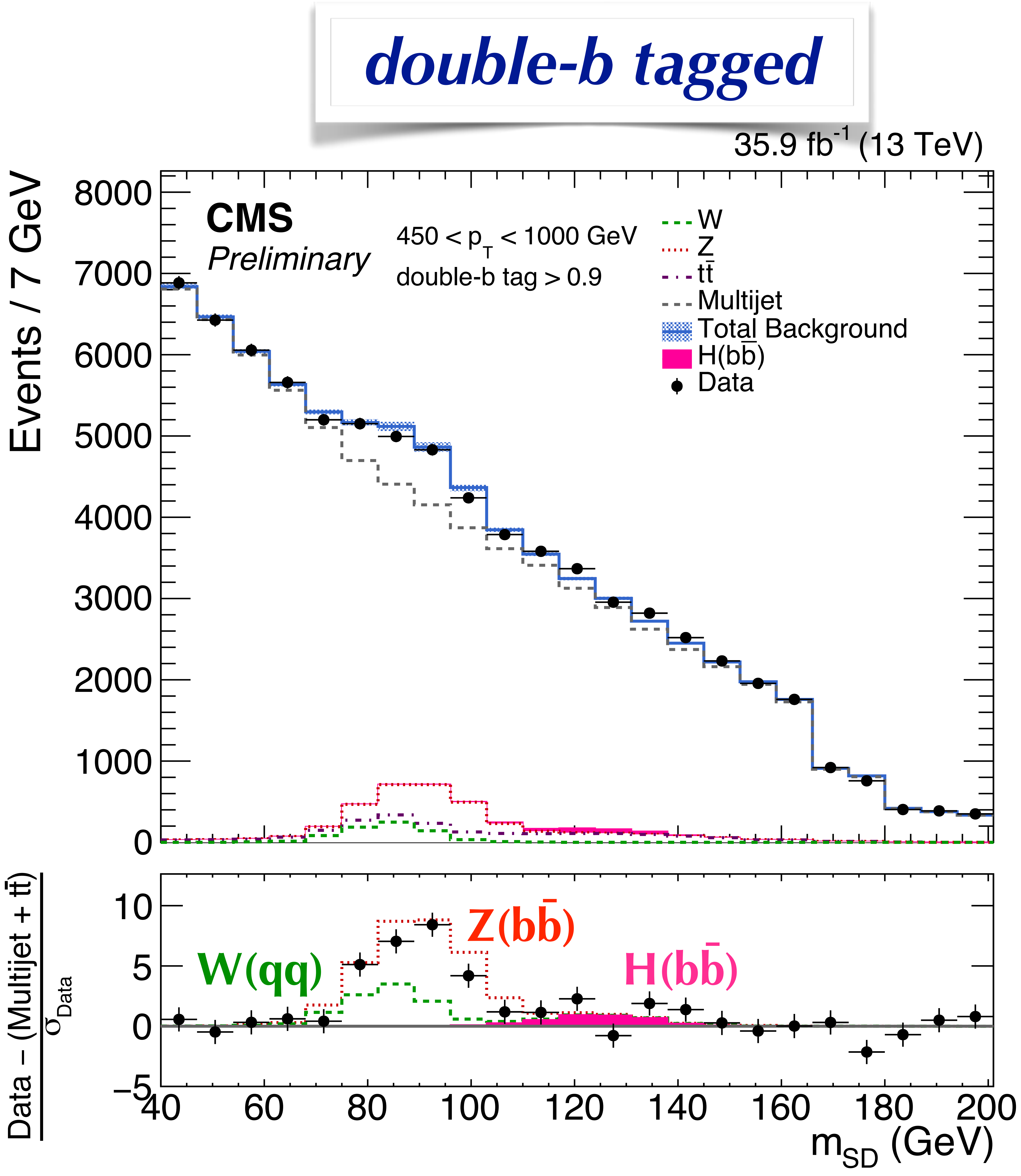
- Observed significance for the Z signal is 5.1σ (5.8 σ expected)
  - **compatible with SM expectation**
- ***First Observation of the Z(b$\bar{b}$) in the one-jet topology***
- This validates the H signal extraction and H-tagging approach



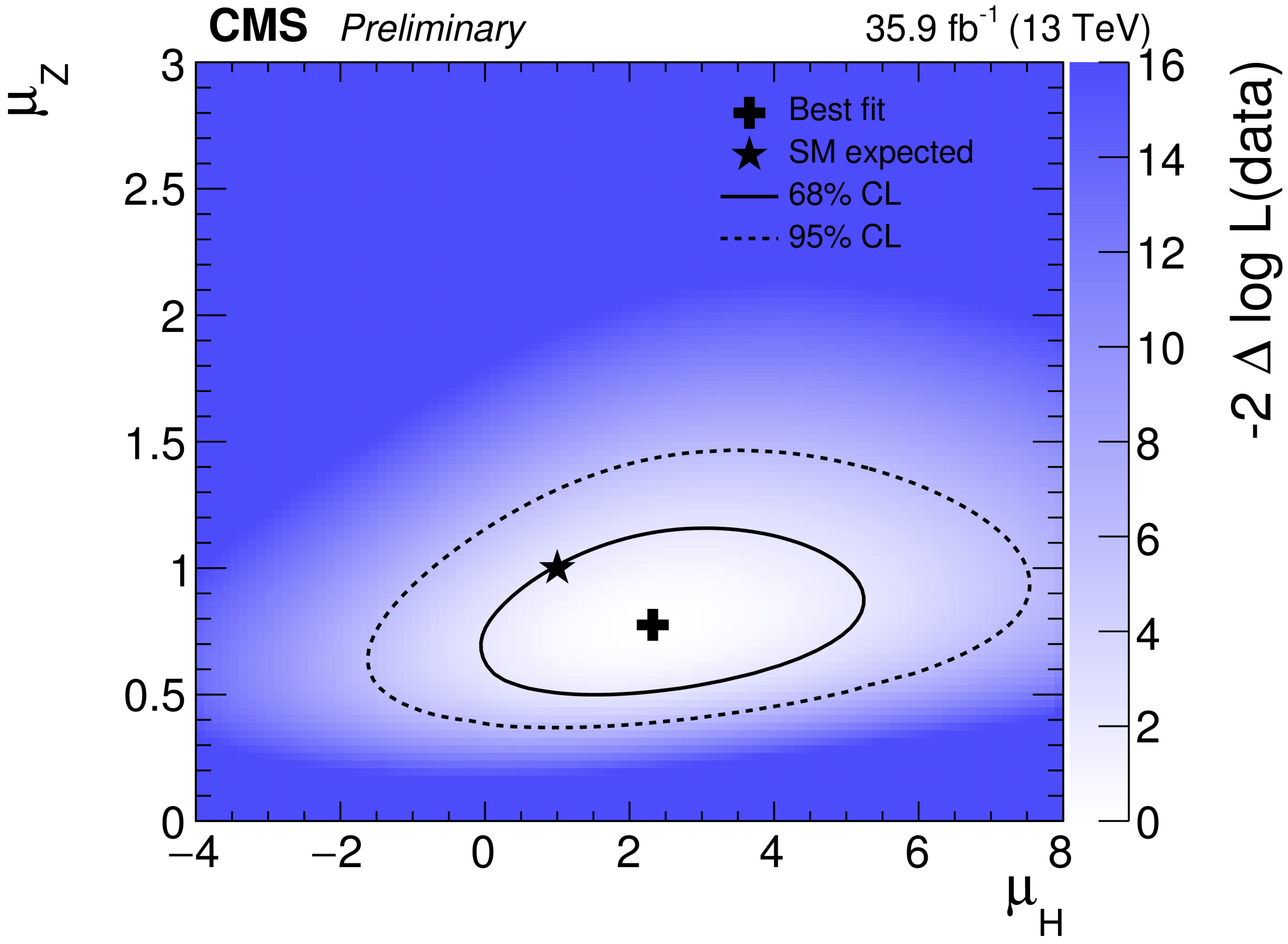
anti-double-b tagged




anti-double-b tagged
CMS
Preliminary
35.9 fb-1 (13 TeV)
450 < pT < 1000 GeV
double-b tag < 0.9
W
Z
tt̄
Multijet
Total Background
H(bb̄)
Data
Events / 7 GeV
Data − (Multijet + tt̄) / σData
W(qq)
Z(qq)
mSD (GeV)


double-b tagged
CMS
Preliminary
35.9 fb-1 (13 TeV)
450 < pT < 1000 GeV
double-b tag > 0.9
W
Z
tt̄
Multijet
Total Background
H(bb̄)
Data
Events / 7 GeV
Data − (Multijet + tt̄) / σData
W(qq)
Z(bb̄)
H(bb̄)
mSD (GeV)

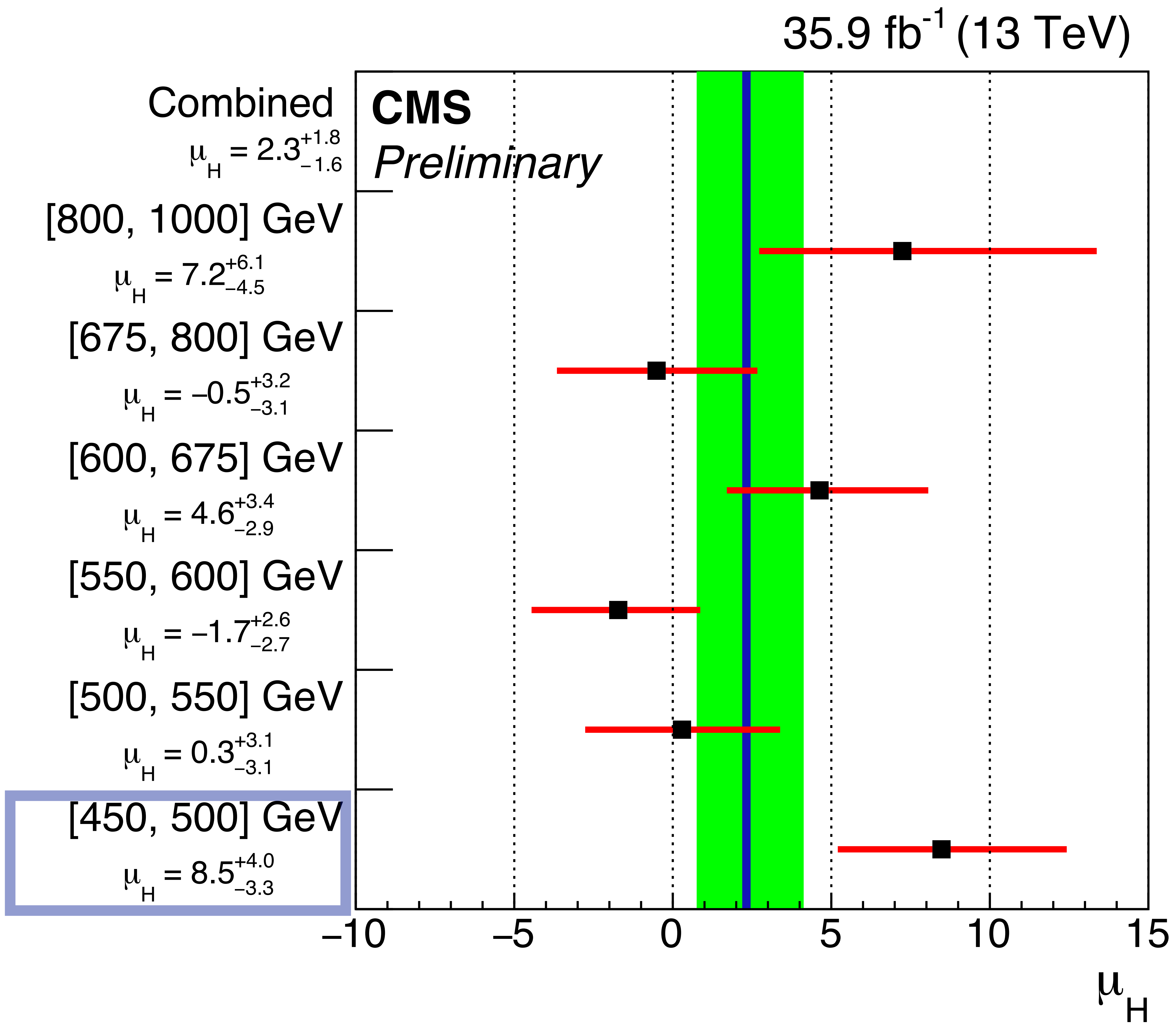
Simultaneous fit of the Z and H signals

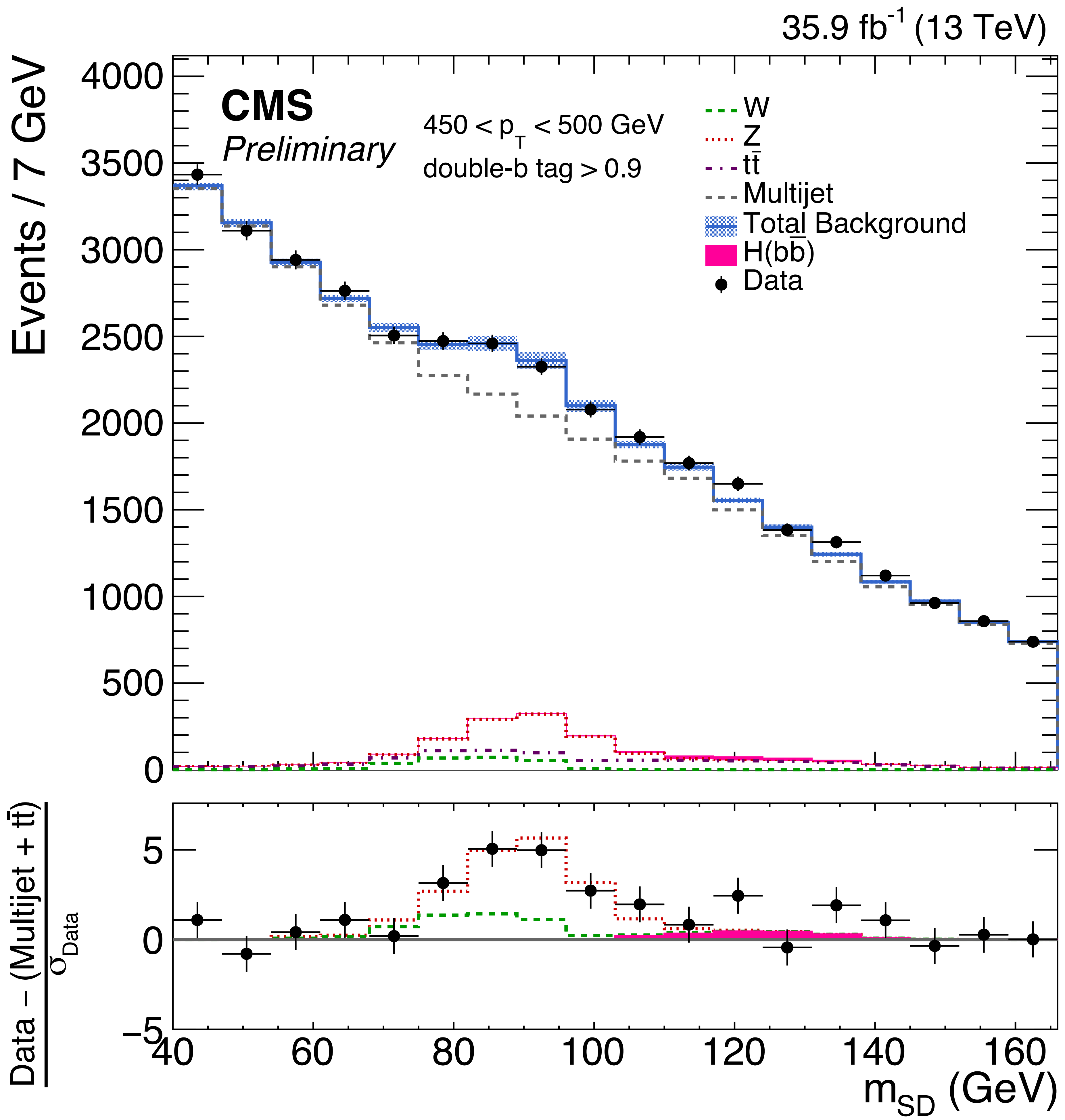


	H	Z
Observed best fit	$\mu_{\mathrm{H}} = 2.3^{+1.8}_{-1.6}$	$\mu_{\mathrm{Z}} = 0.78^{+0.23}_{-0.19}$
Expected significance	0.7σ ($\mu_{\mathrm{H}} = 1$)	5.8σ ($\mu_{\mathrm{Z}} = 1$)
Observed significance	1.5σ	5.1σ

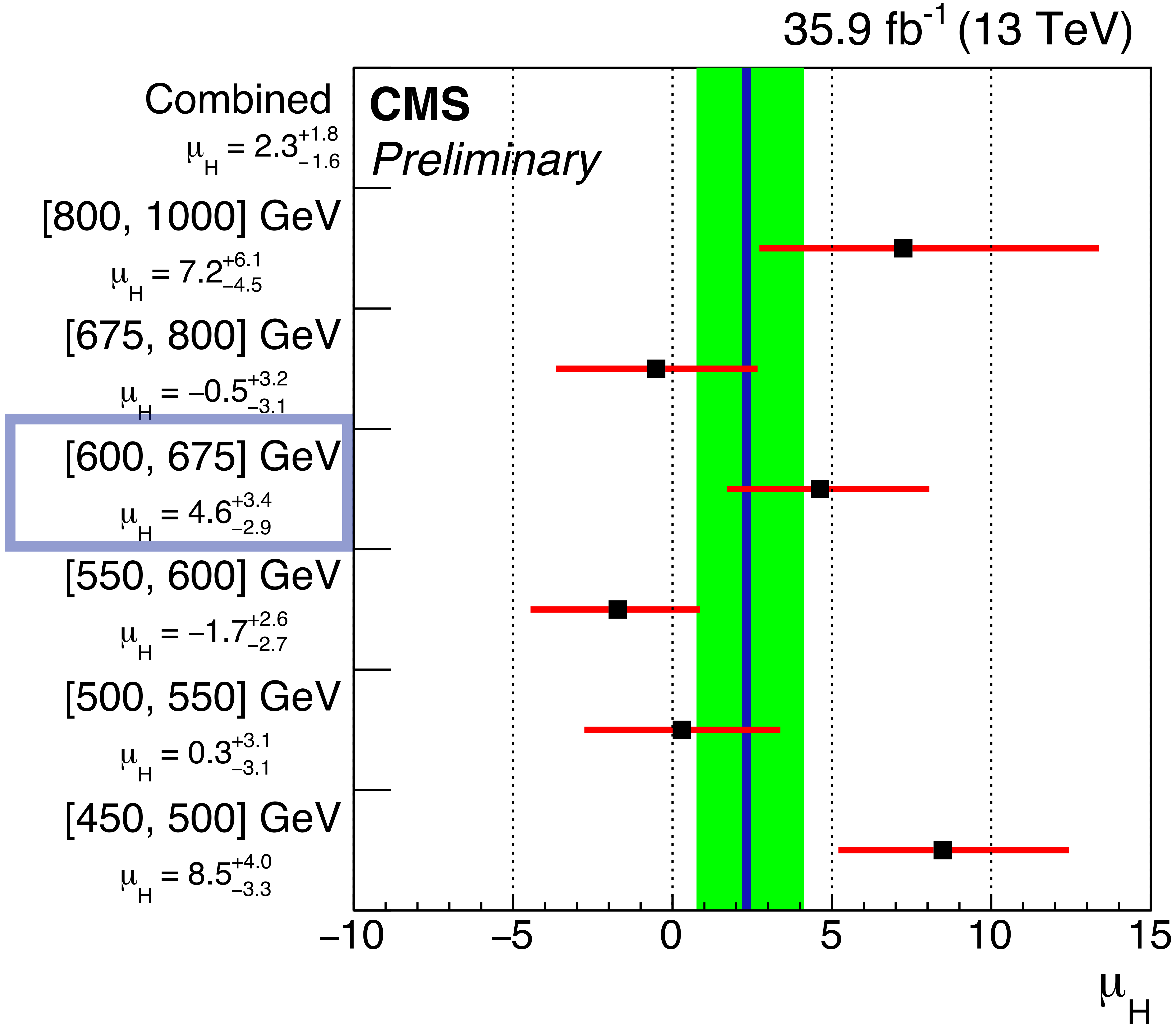
*Observed significance for the Z and H signals is **compatible with SM***

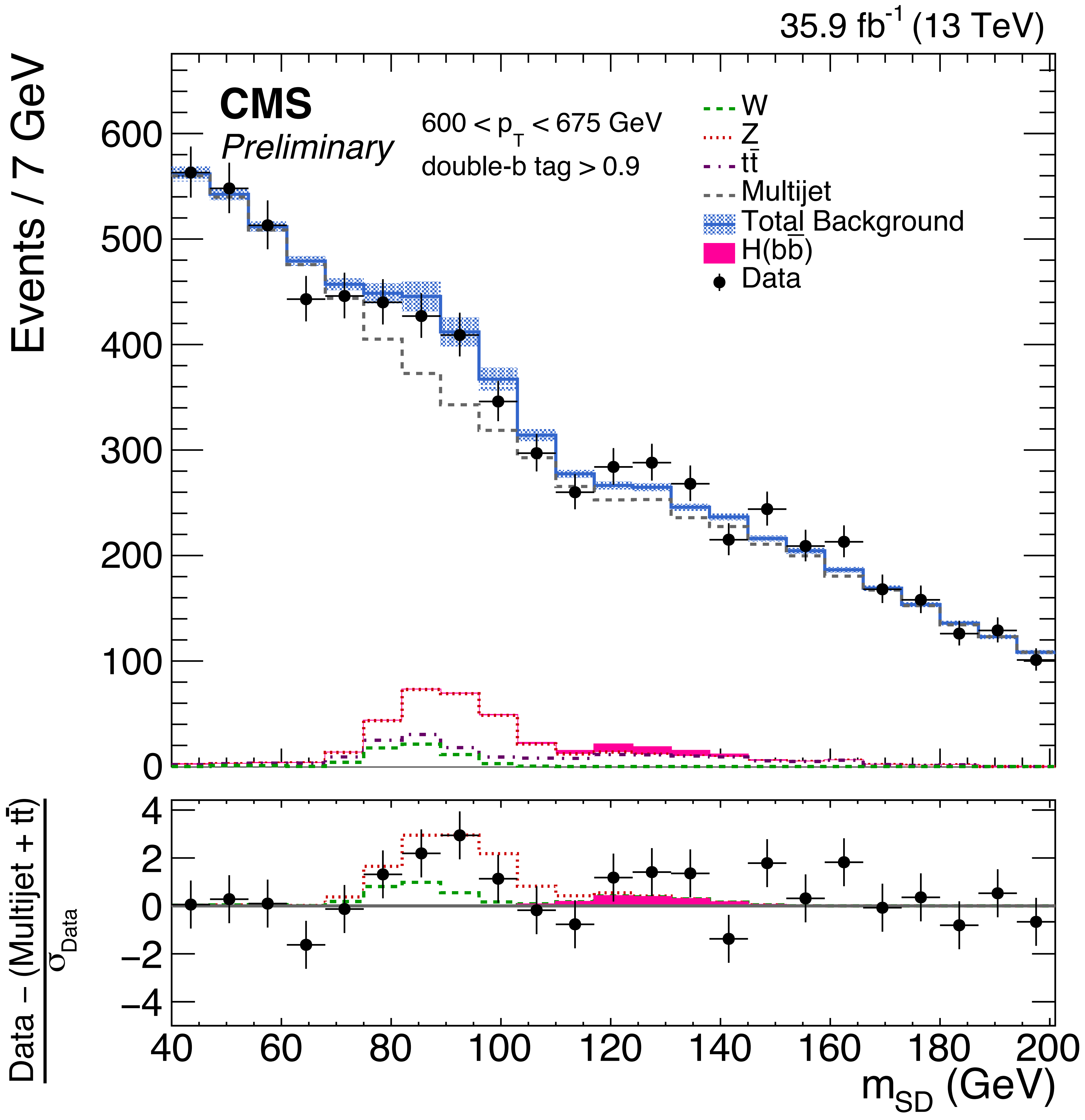
Sensitivity per p_T category



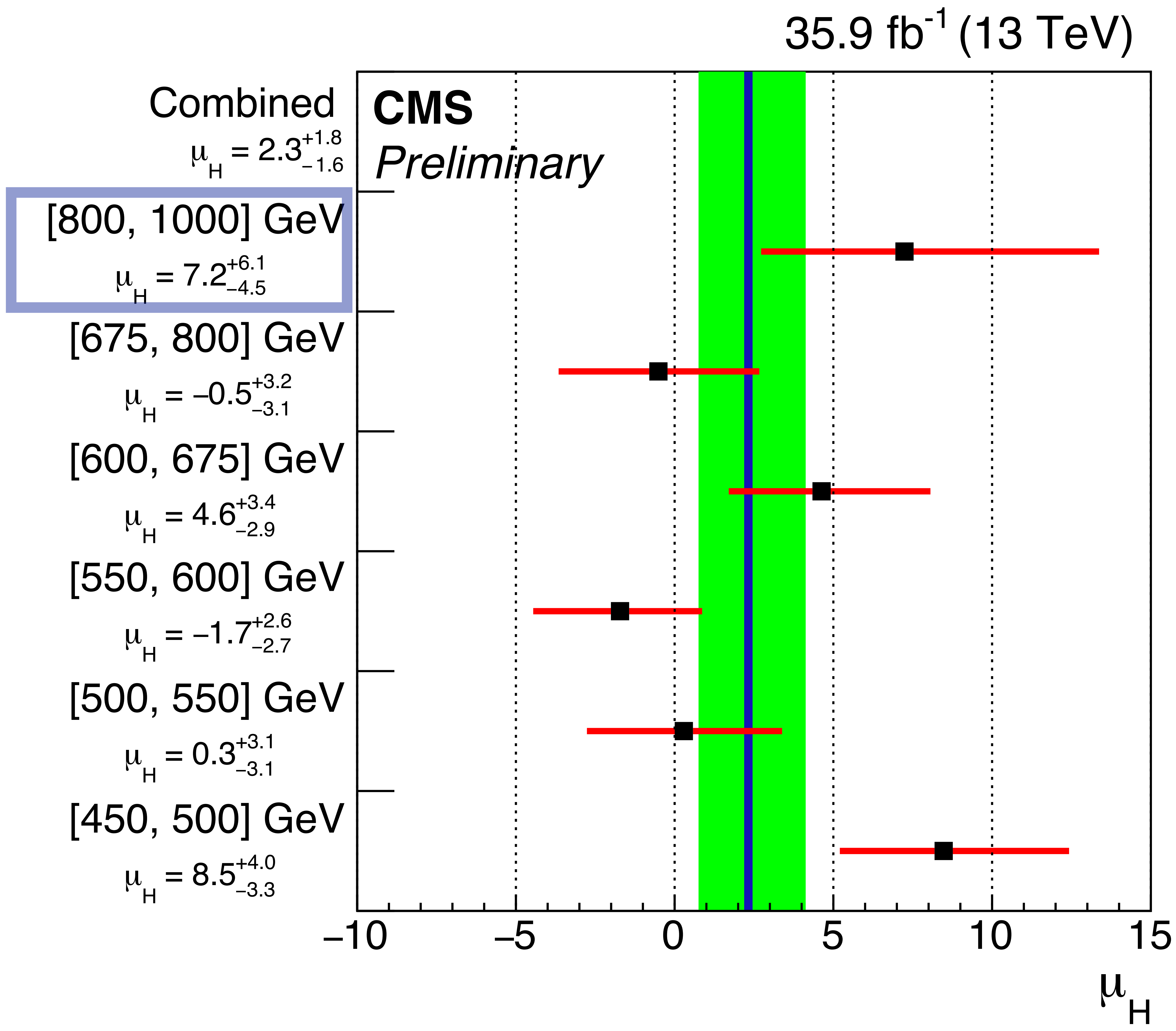


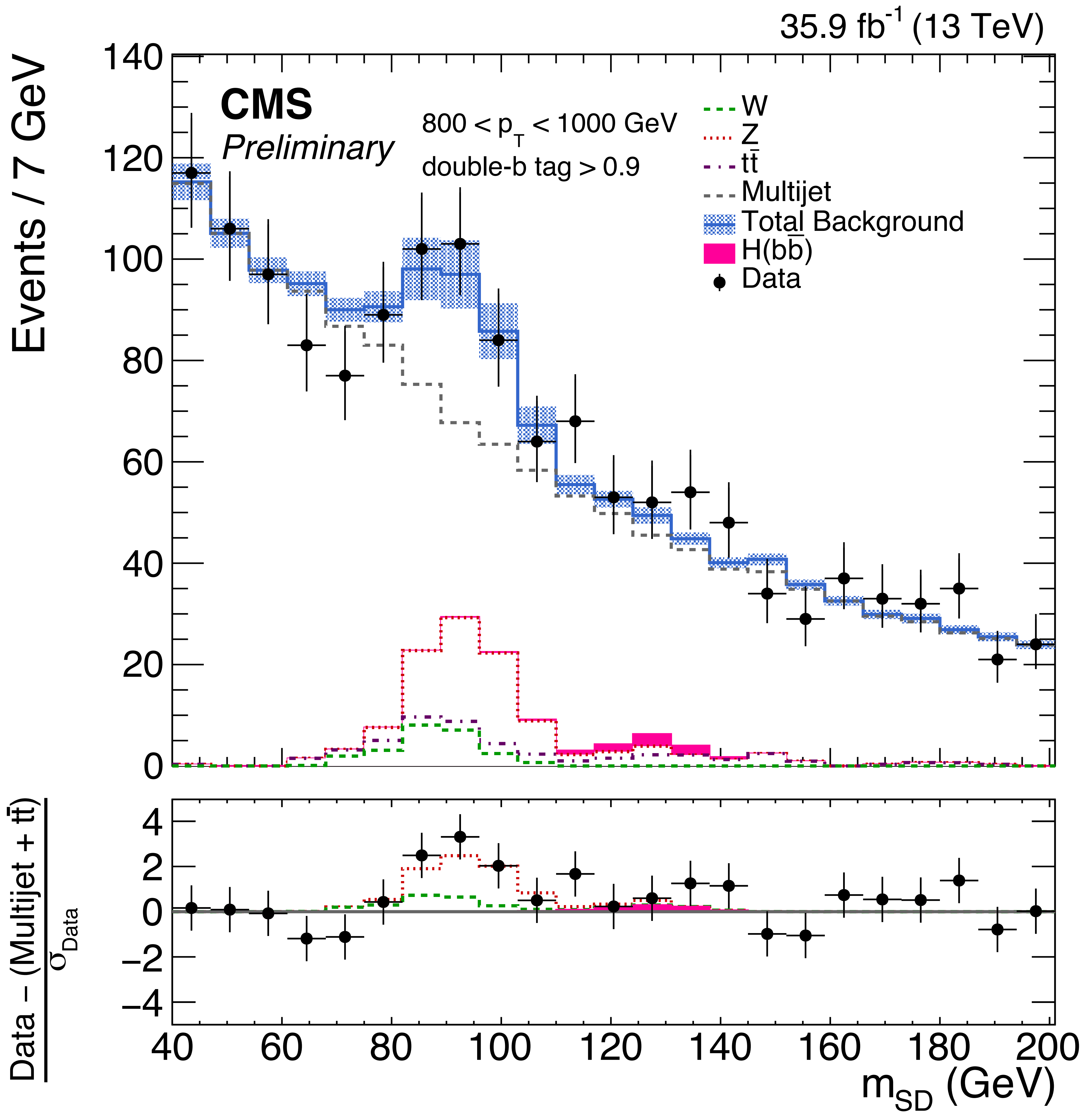
Sensitivity per p_T category





Sensitivity per p_T category





Measured cross section

- The measured cross sections for Z+jets and Higgs for jet p_T > 450 GeV are:

$$\sigma_Z = 849\ {+257/{-209}}\ \text{fb}$$
$$\sigma_H = 74\ {+51/{-49}}\ \text{fb}$$

- Broken down into:

$$\sigma_Z = 849\ {+155/{-155}}\ (\text{stat.})\ {+140/{-205}}\ (\text{syst.})$$
$$\sigma_H = 74\ {+48/{-48}}\ (\text{stat.})\ \ {+10/{-17}}\ (\text{syst.})$$

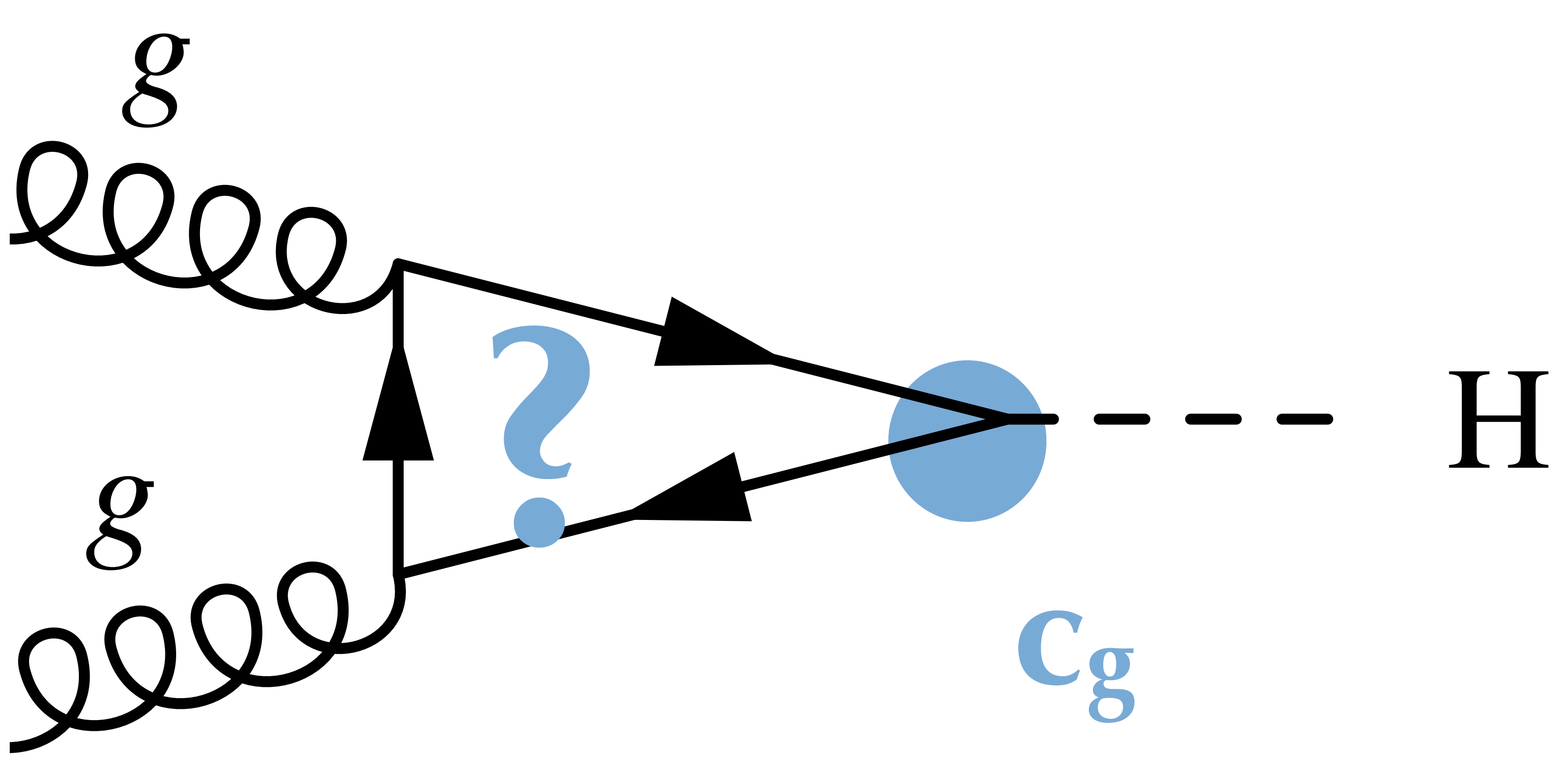
Compatible with SM within uncertainties

Outline

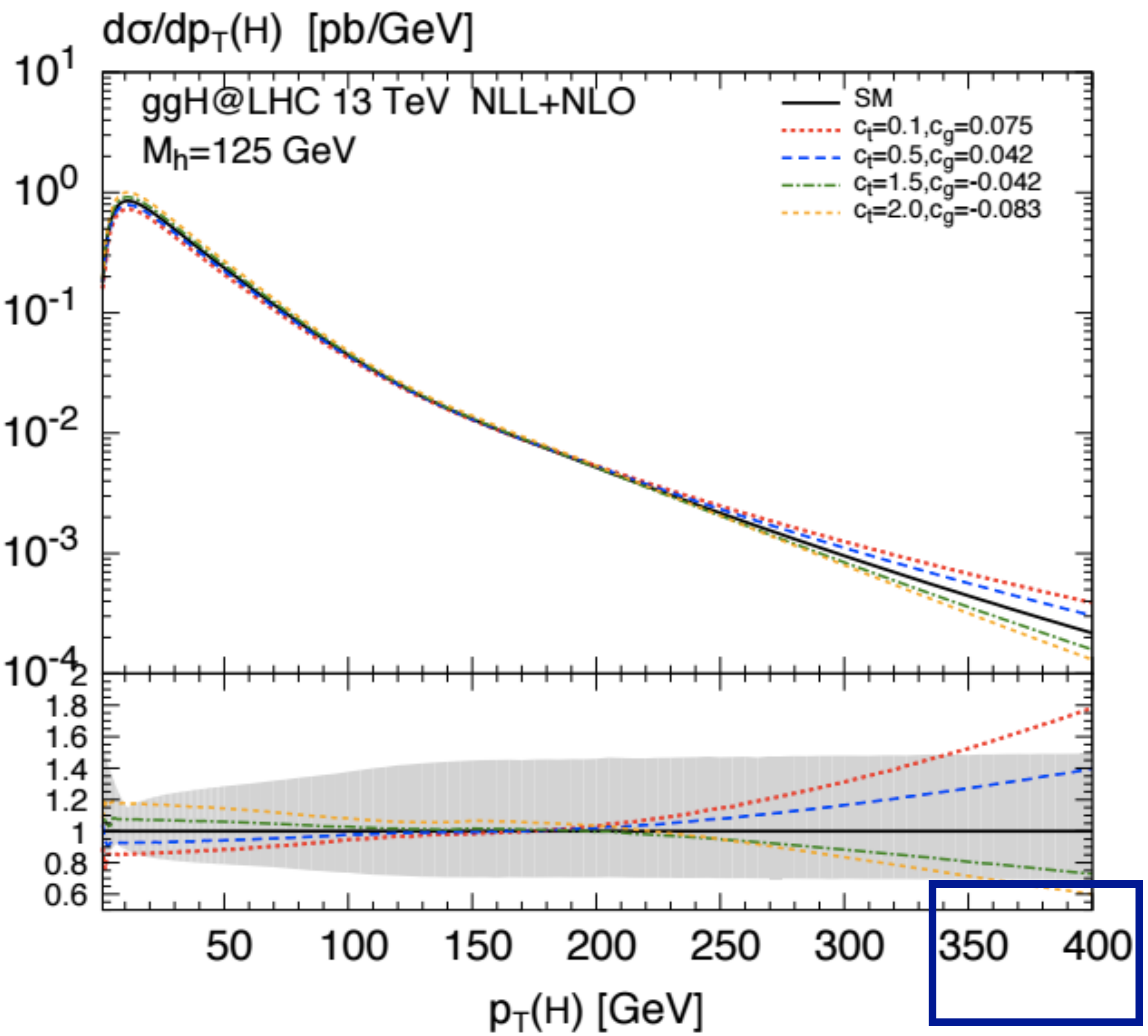
- Higgs boson discovery at LHC
- $H \to b\bar{b}$ state of the art
- Tools for identifying $H \to b\bar{b}$ at high p_T
  - b-quark identification in CMS
- Inclusive search for boosted $H \to b\bar{b}$
- **Future perspectives**

ArXiv:1612.00283
JHEP10(2016)123

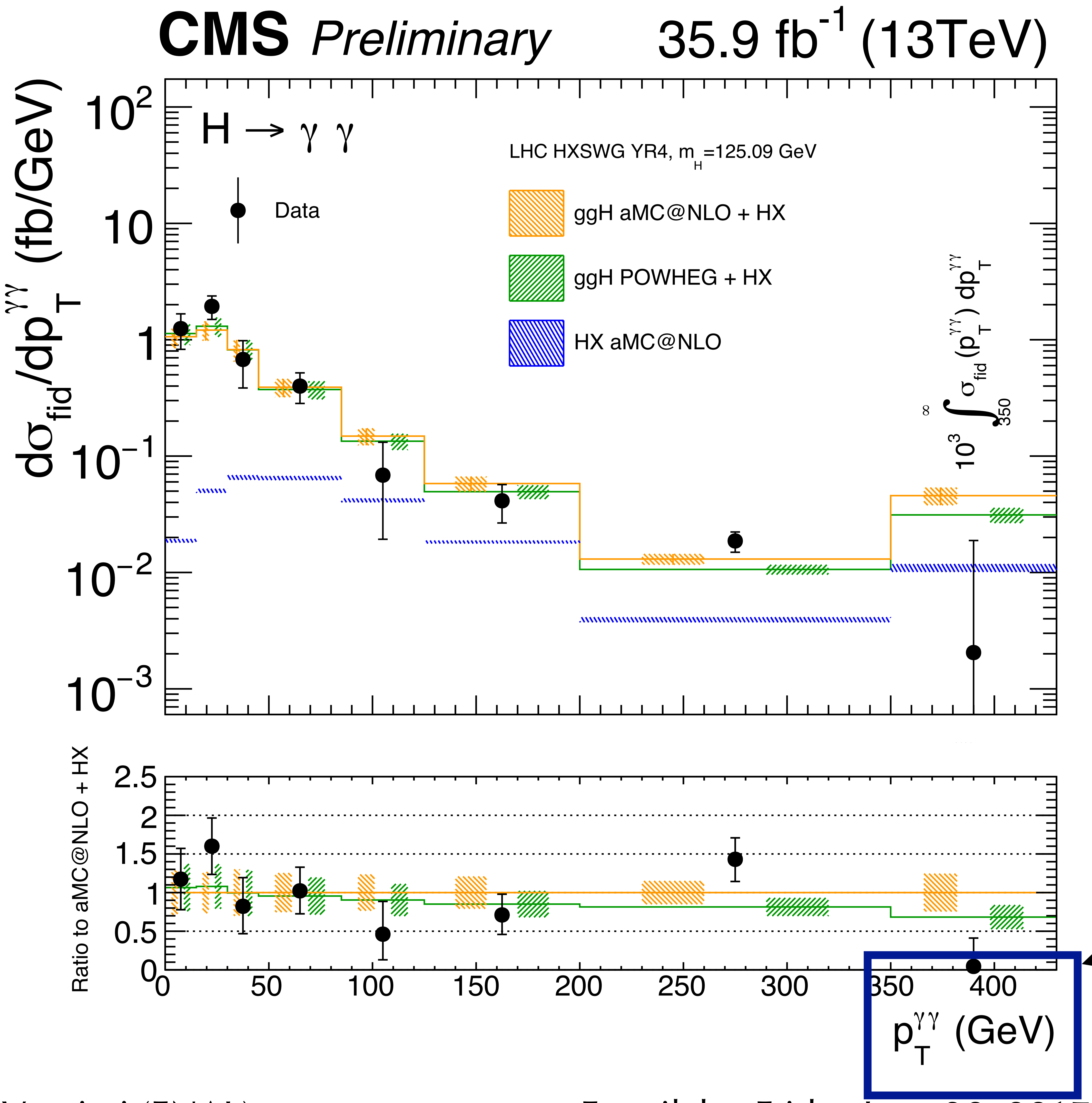
Sensitivity to BSM

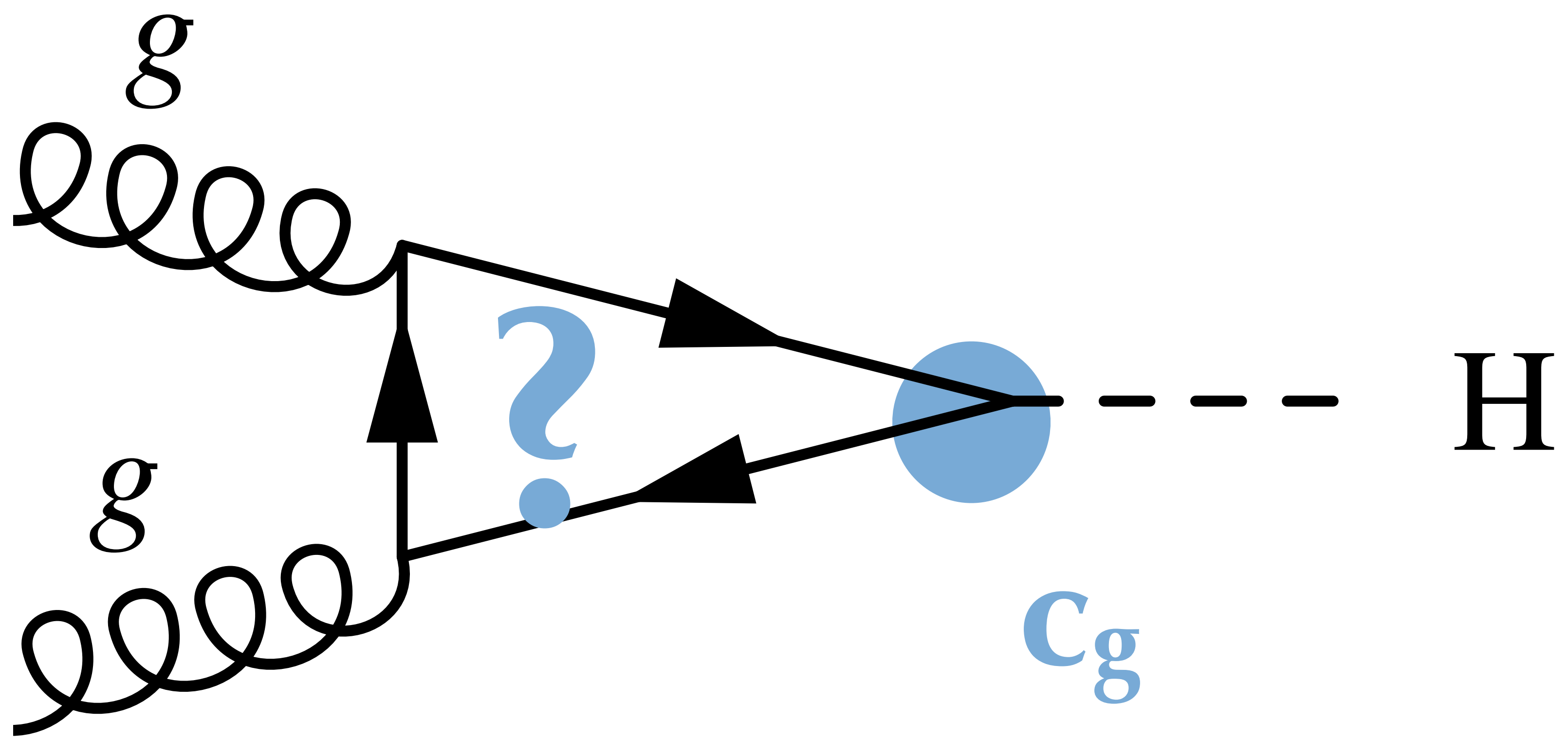


- Probing **gluon coupling vs top coupling**
- At high p_T effective vertex dominated by top quark
  - Directly probe **modifications in top quark coupling**



Current Higgs p_T measurement





H(b$\bar{b}$), this analysis, starts here

Conclusions and perspectives

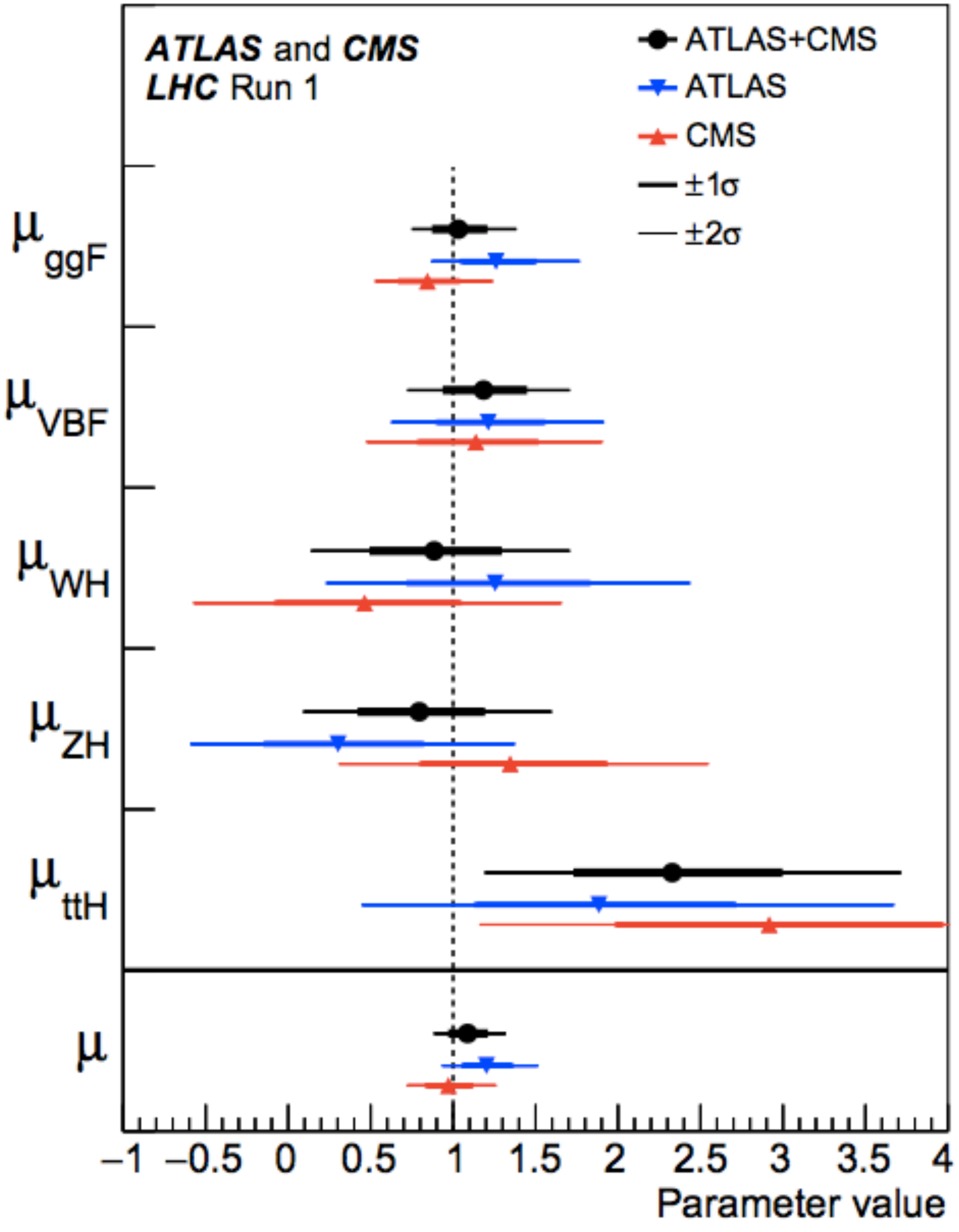
- **First search** for $gg \rightarrow H \rightarrow b\bar{b}$ in boosted topology
  - **First observation of $Z(b\bar{b})$** in single jet topology, **5.1σ**
  - The observed significance for the **$H(b\bar{b})$** is **1.5σ**
  - Cross sections are measured and agree with SM
- This search looks at previously **unexplored regions of phase space** and opens a **new strategy** to:
  - search for Higgs boson to $b\bar{b}$
  - probe **BSM contributions to the Higgs at very high p_T**
  - investigate **new physics final states** involving high p_T Higgs boson

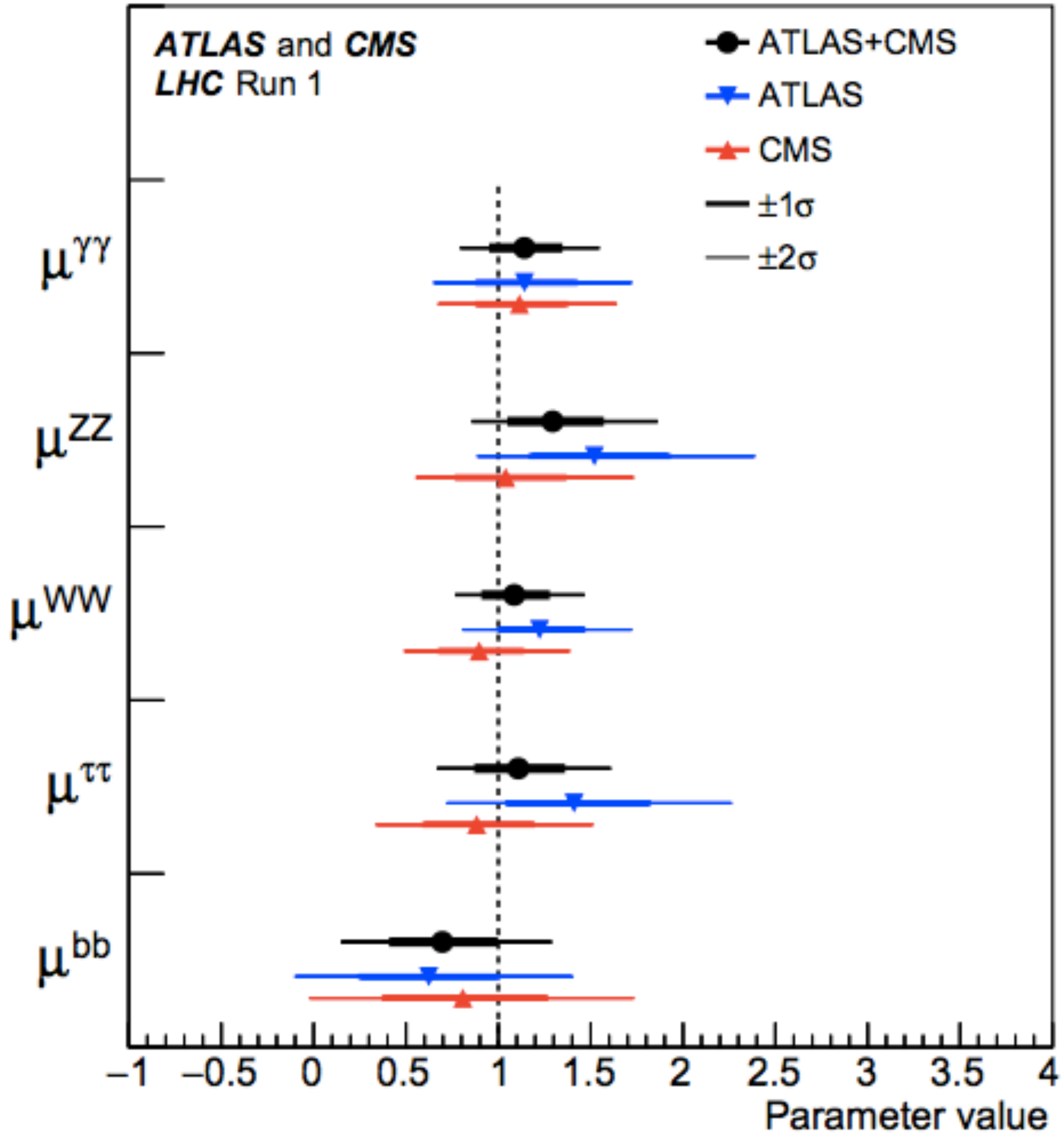
thank you!

–Additional Material

JHEP08(2016)045

Higgs at LHC

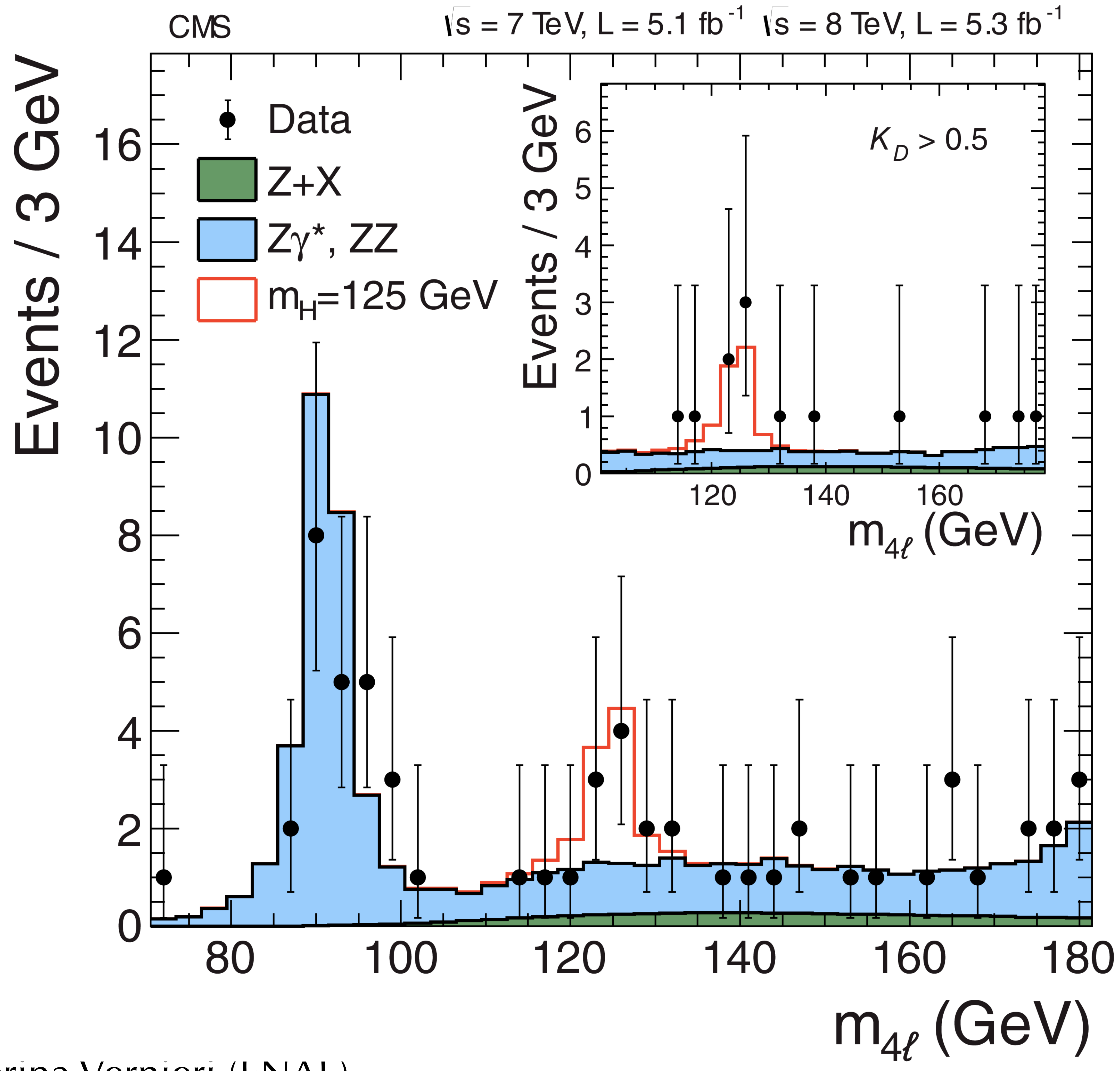




Phys. Lett. B 716 (2012) 30
Phys. Rev. D 89 (2014)

Challenges of the H($b\bar{b}$) mode at the LHC

Comparison with one of the discovery channels



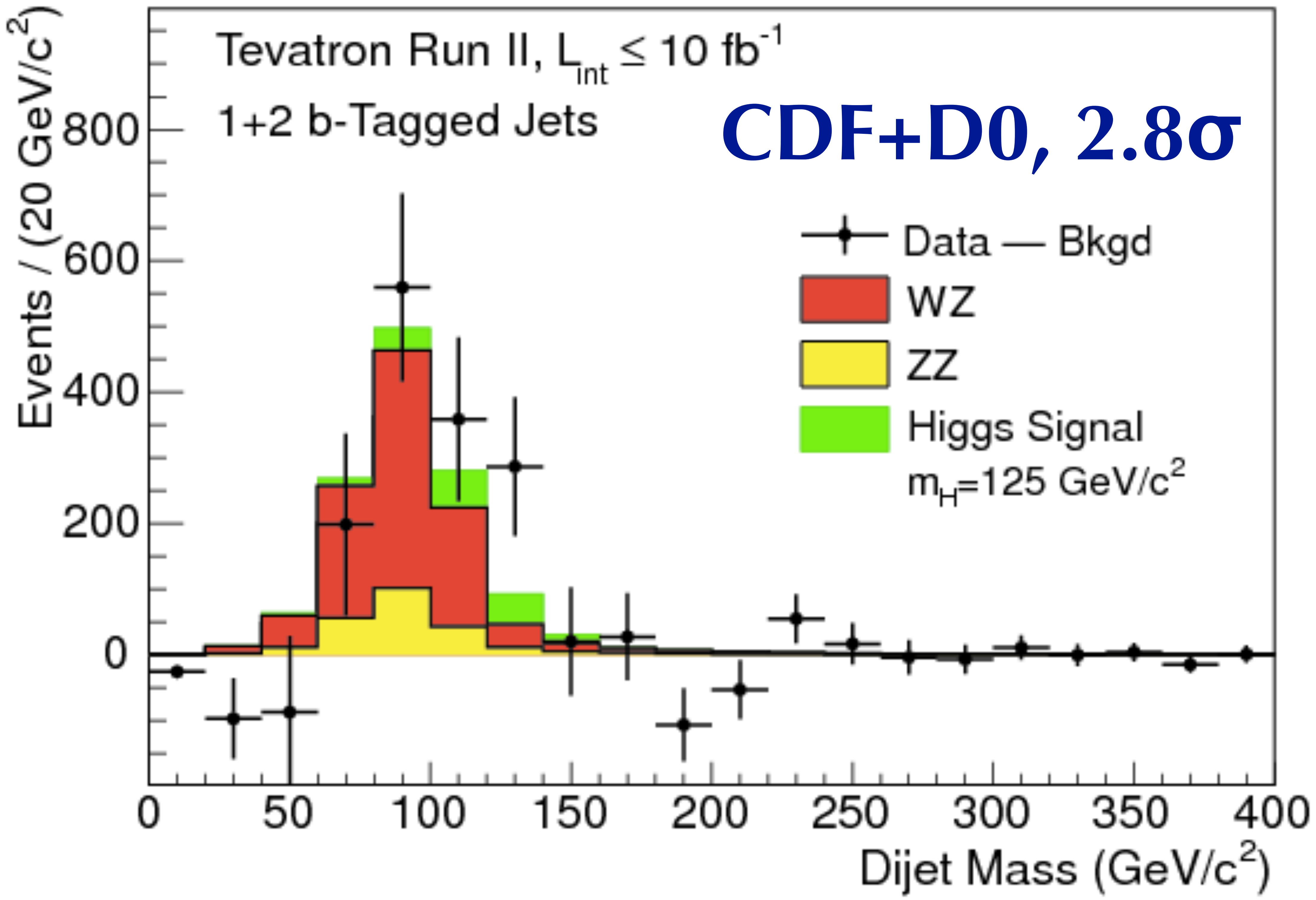
	$H \to 4\ell$	$H \to b\bar{b}$
BR	0.03%	58%
mass resolution	1%	10%
signal efficiency	30%	1.3%
S/B	2	0.05

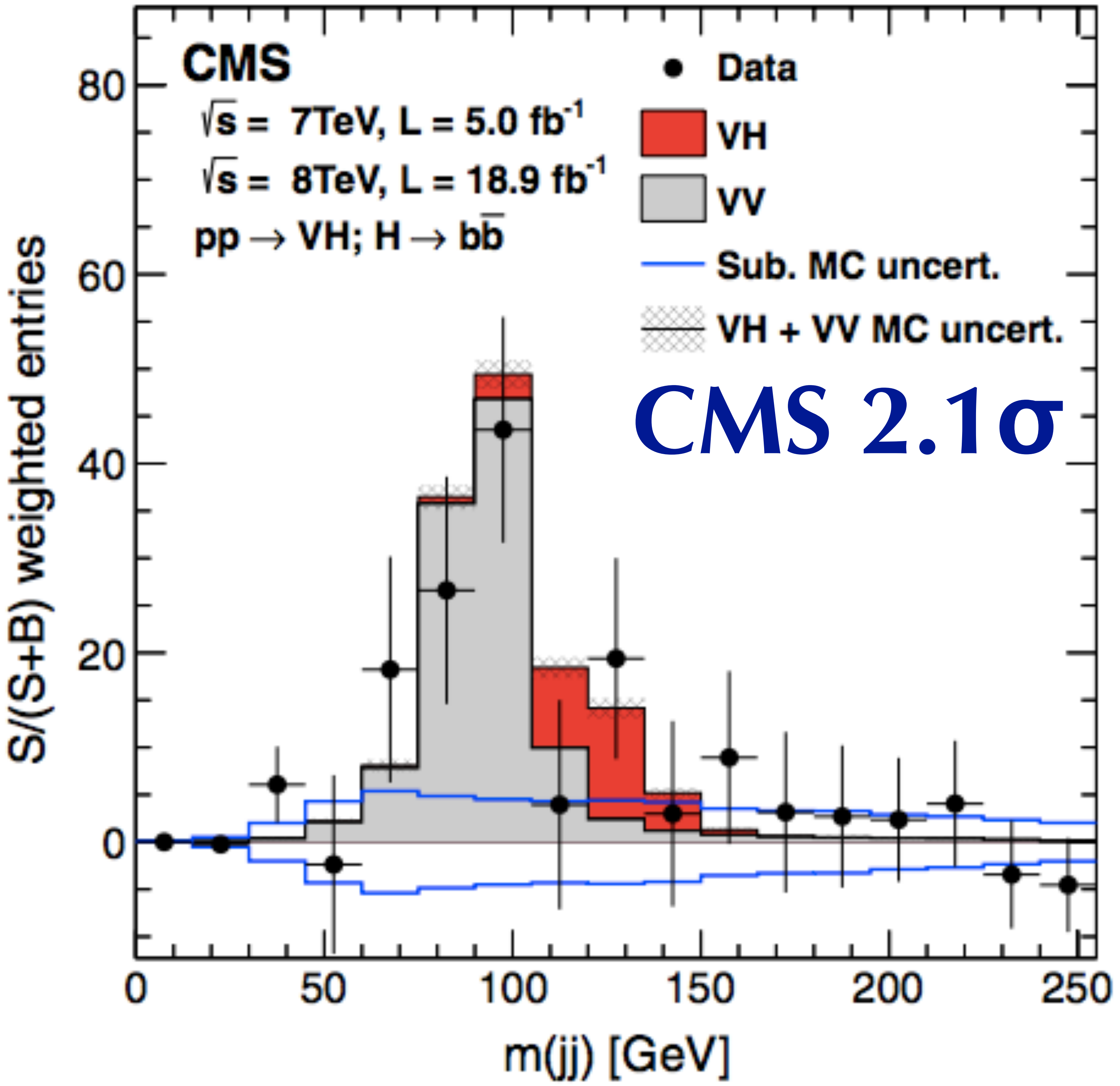
<u>H($b\bar{b}$) searches need:</u>

- b-jets identification
- improve m($b\bar{b}$) resolution
- exploit all possible information from the event to improve S/B

$VH(b\bar{b})$

Phys. Rev. D 89 (2014)
Phys. Rev. Lett. 109 071804

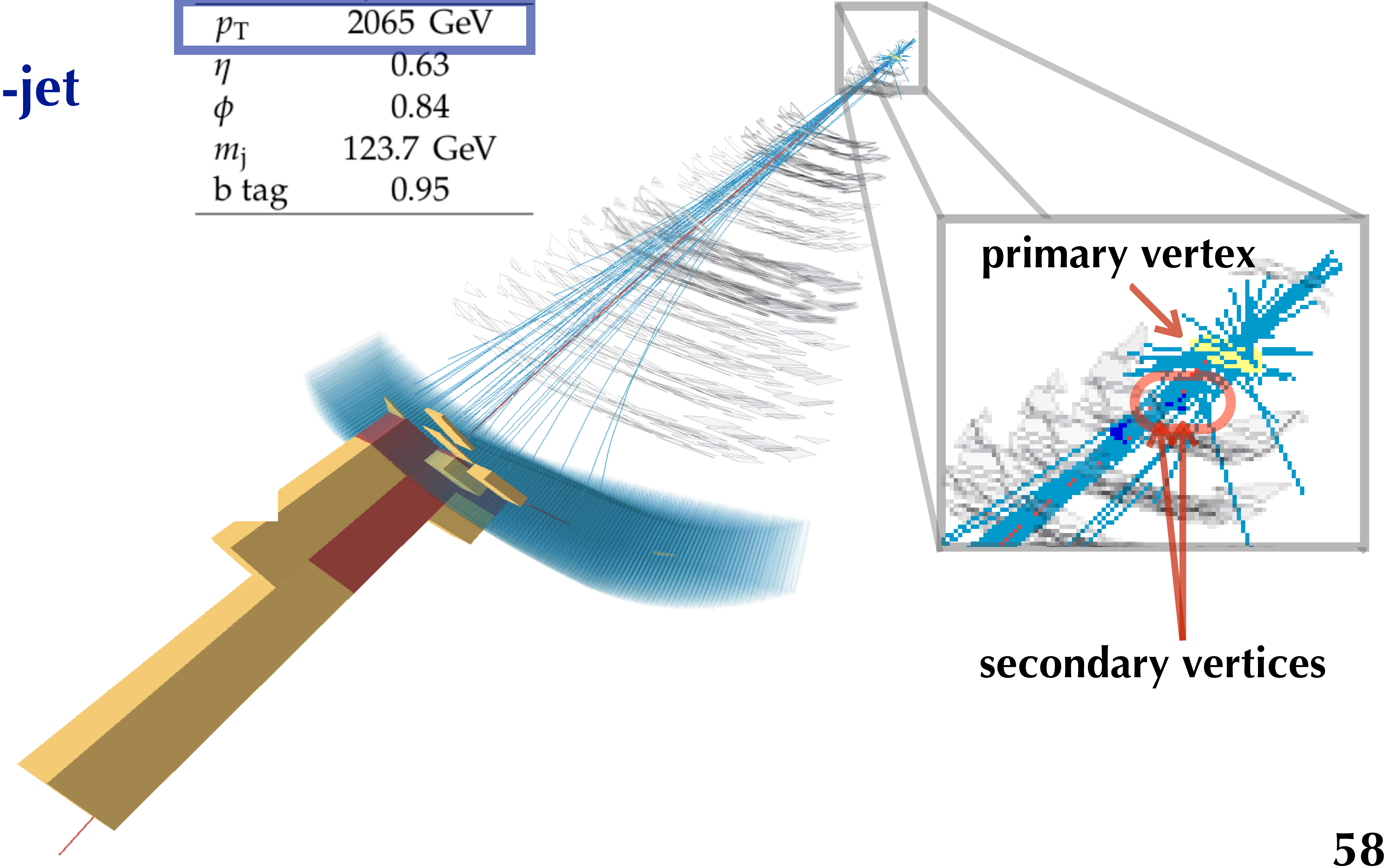




S/B at LHC is 2.5x lower than at Tevatron

H($b\bar{b}$)-jet

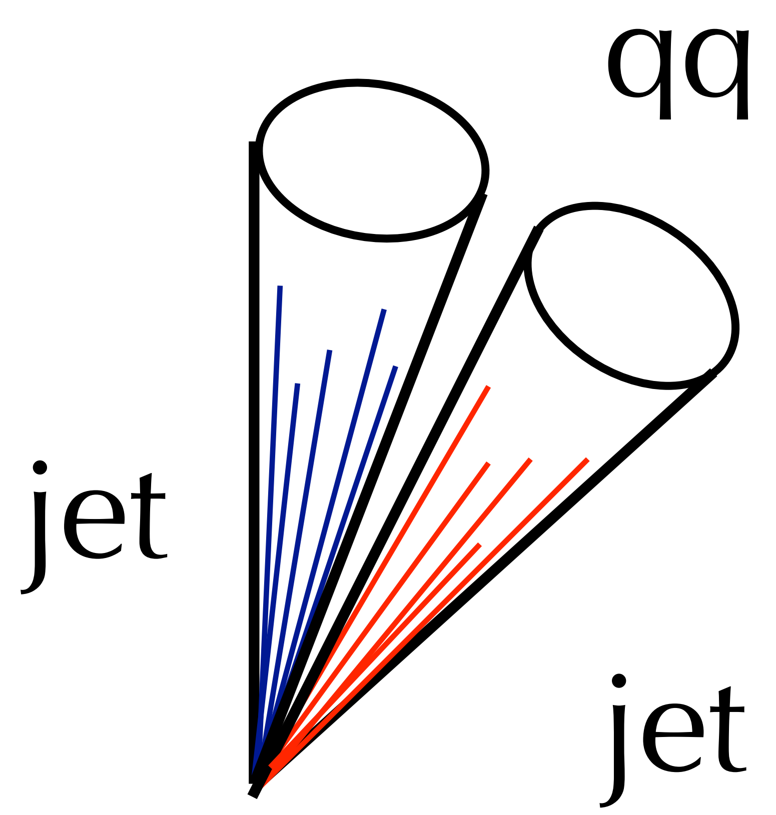
p_T	2065 GeV
η	0.63
ϕ	0.84
m_j	123.7 GeV
b tag	0.95

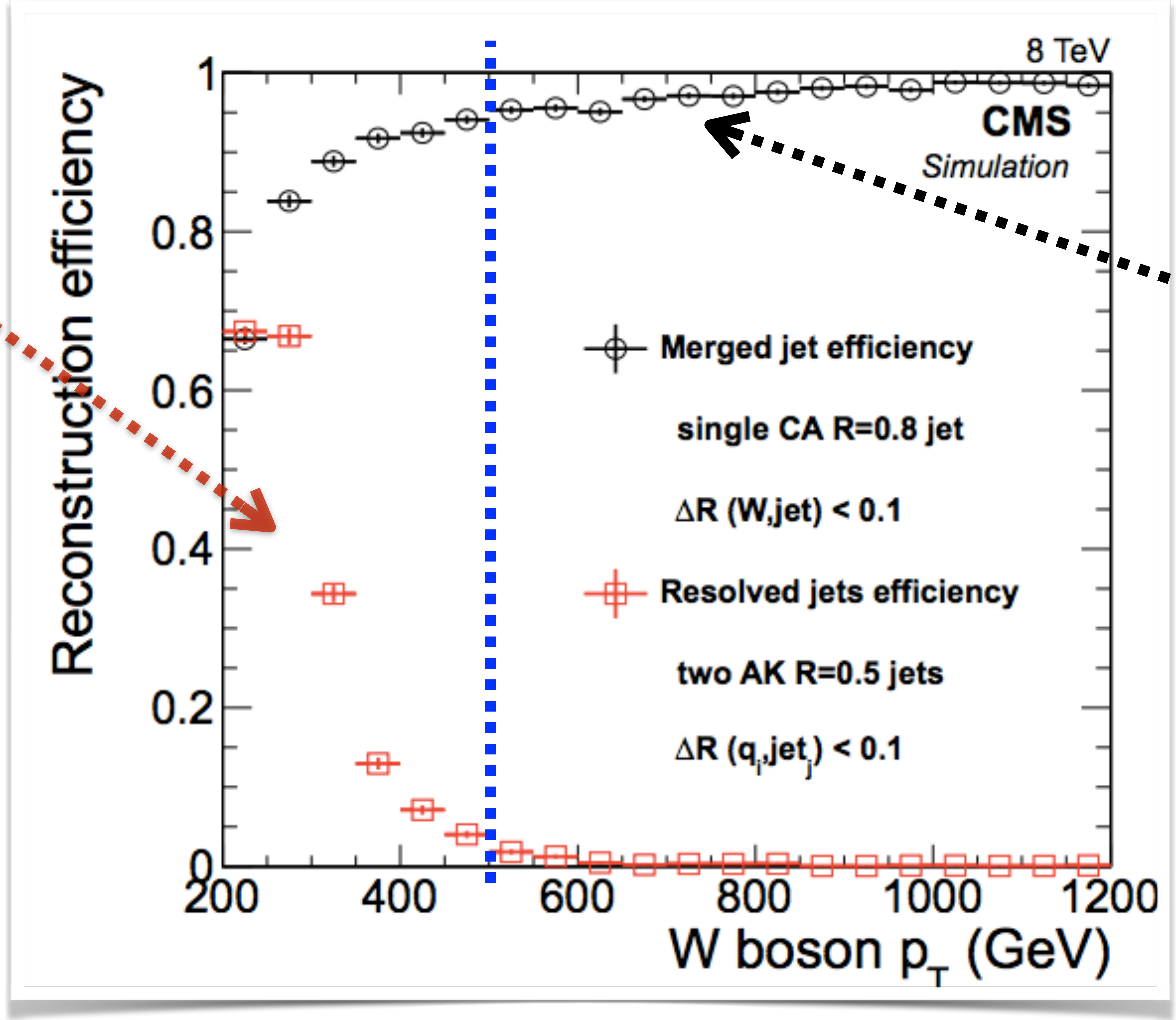


Boosted bosons

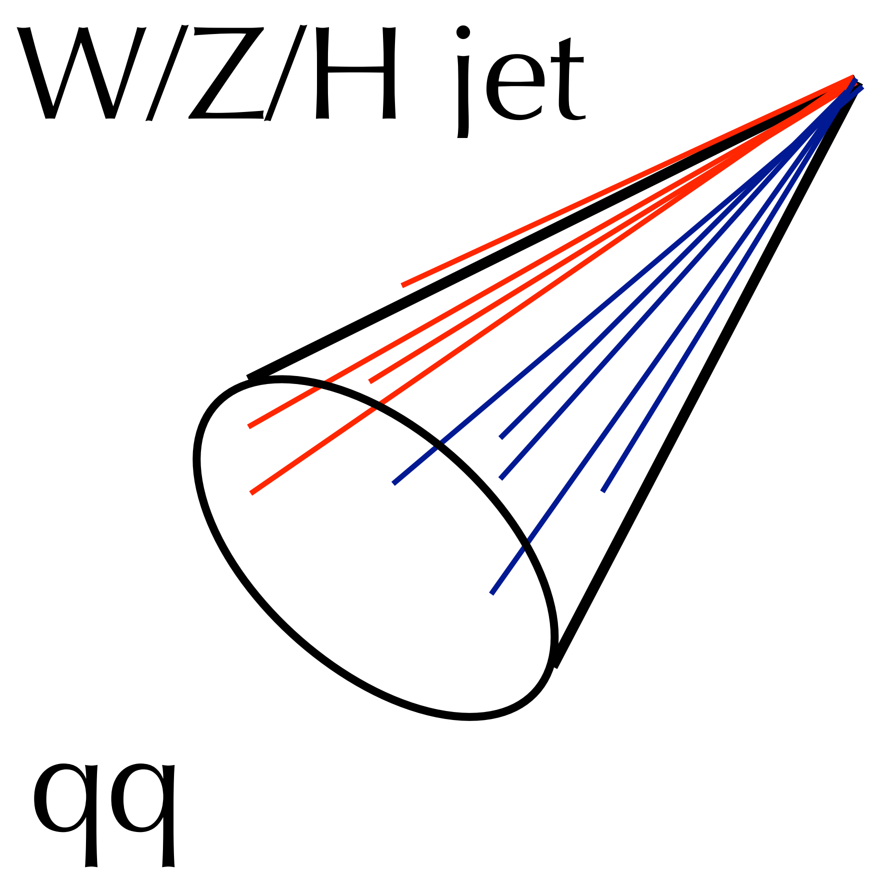
Bosons produced with high p_T merge into single large-R jet (0.8 CMS, 1.0 ATLAS)

two-separated jets



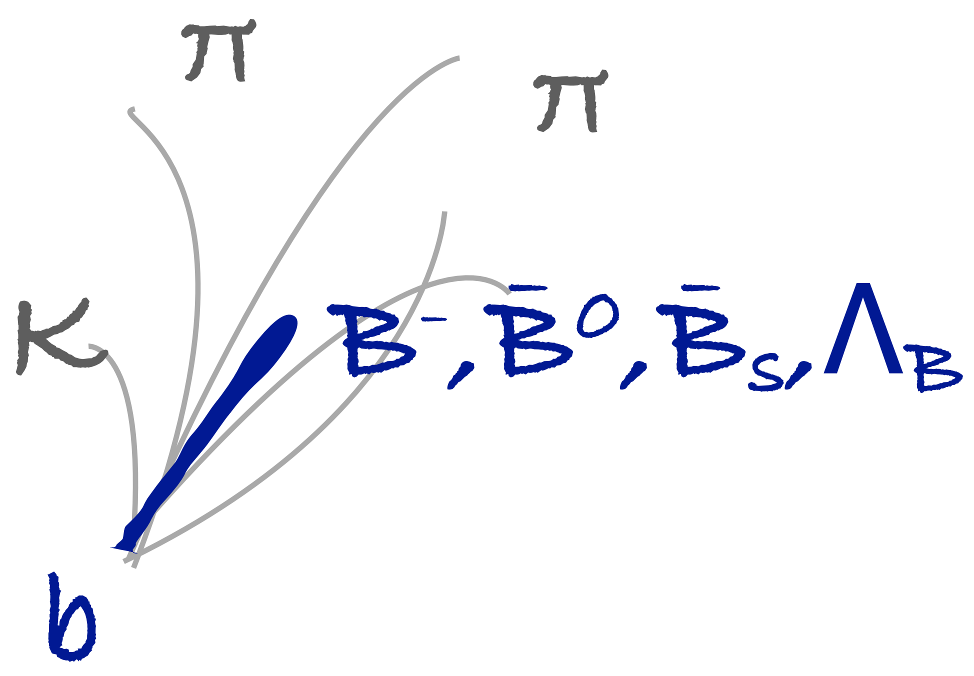


one single large-cone (fat) jet



B properties useful for its tagging

b quarks hadronize

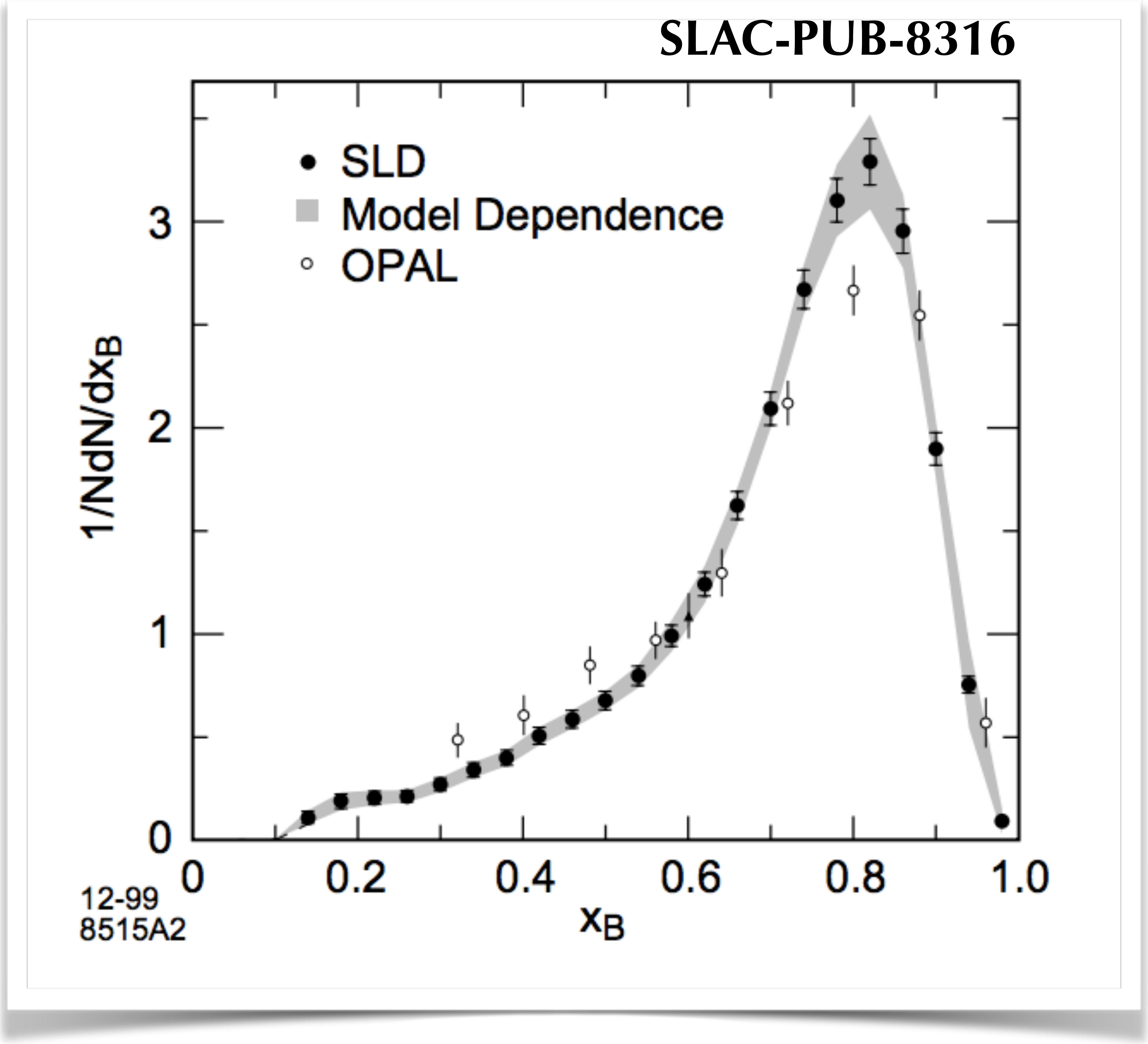


- **Measurable lifetime**

  $c\tau \sim 500\ \mu m \rightarrow \beta\gamma c\tau \sim 5mm$ @ 50 GeV
- **Large mass** (~5 GeV)
- The **weak b-decay** often produces leptons

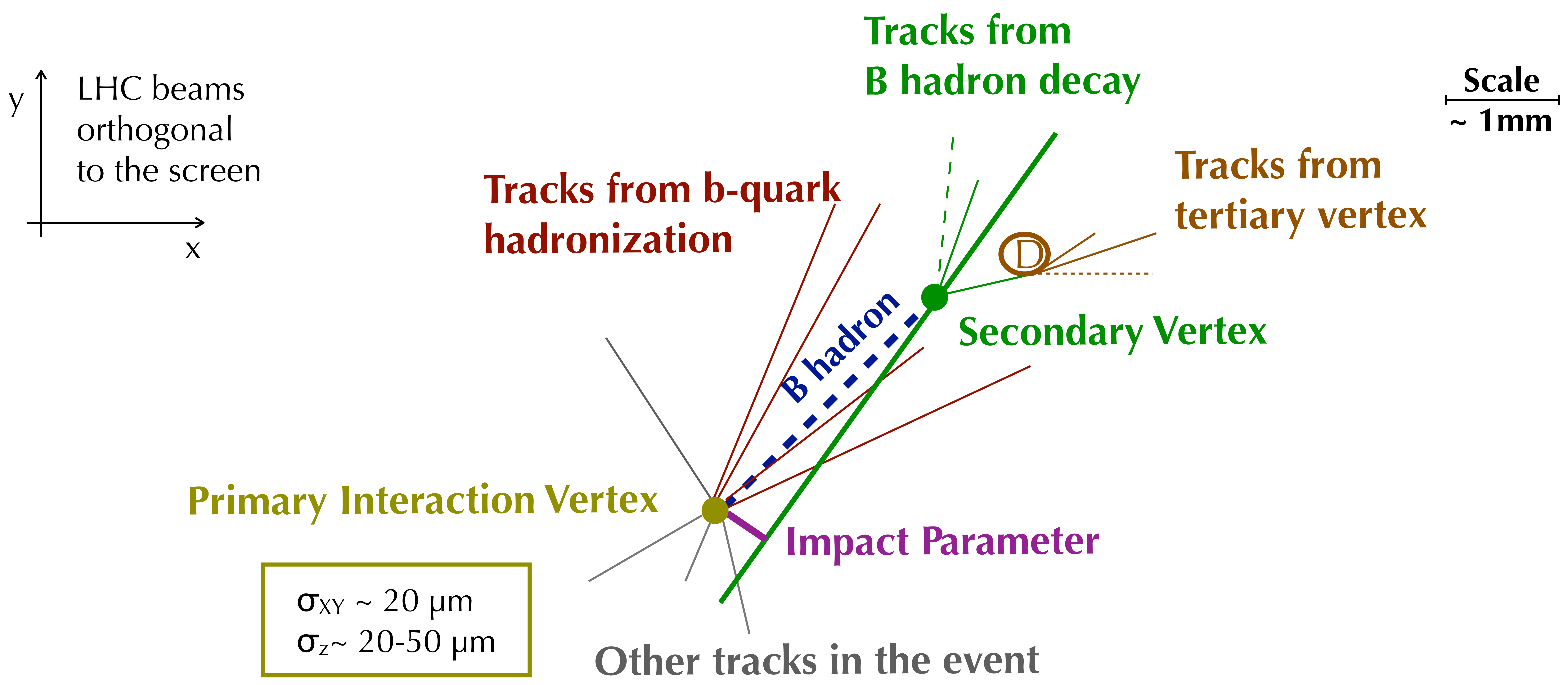
  BR: $B \rightarrow l + \nu + X$ ~ 25%

  $B \rightarrow D \rightarrow l + \nu + X'$ ~ 20% **tertiary vertex**
- **High momentum transferred** to the B hadron



Fraction of the original b-quark momentum carried by the B
$\langle x_B \rangle \sim 0.7$

The B tag picture



σ_{XY} ~ 20 μm
σ_{z}~ 20-50 μm

Combined algorithms

The **Combined Secondary Vertex** through multivariate technique combines (CSVv2)

Track information

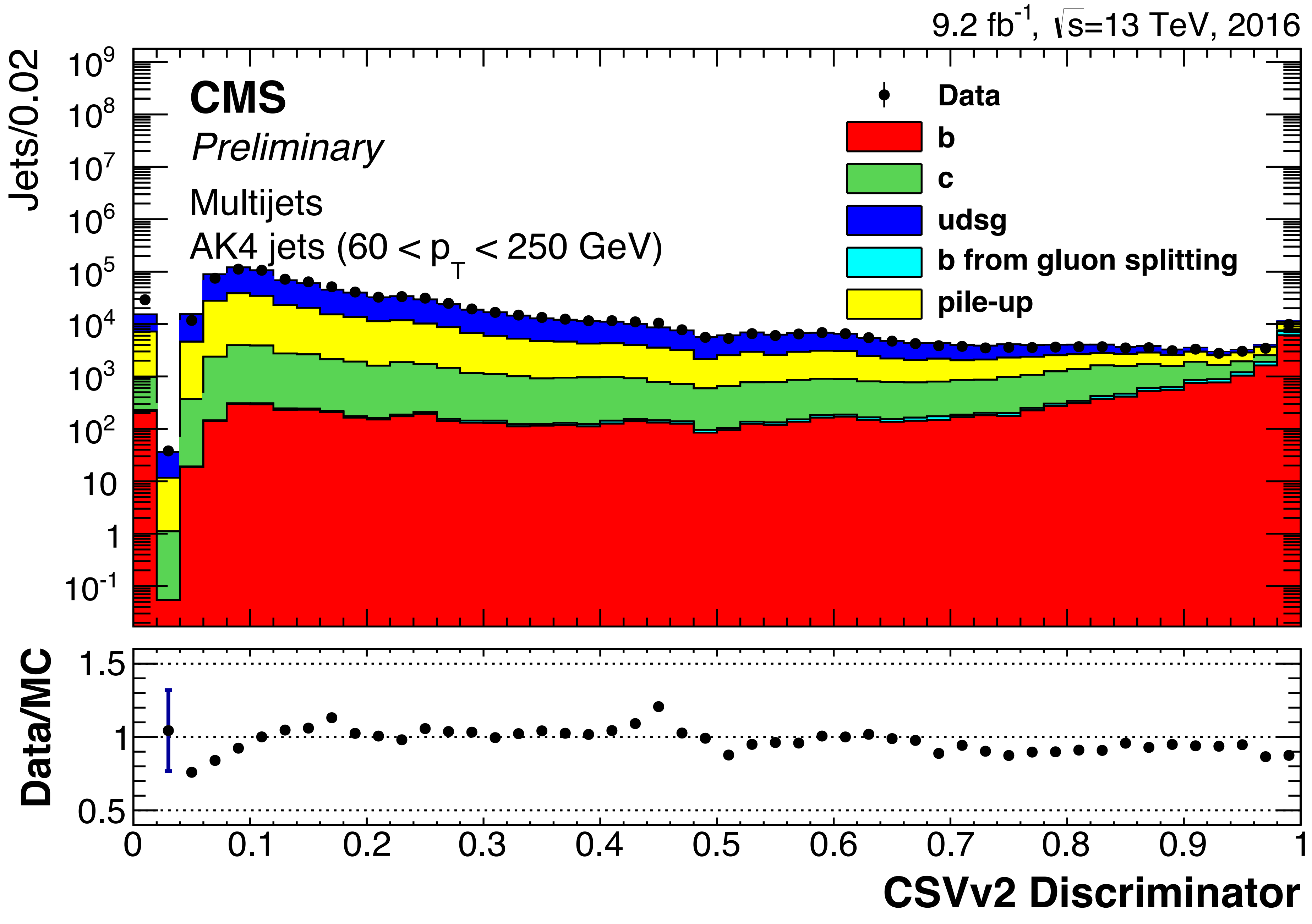
- 3D IP significance of the most energetic tracks

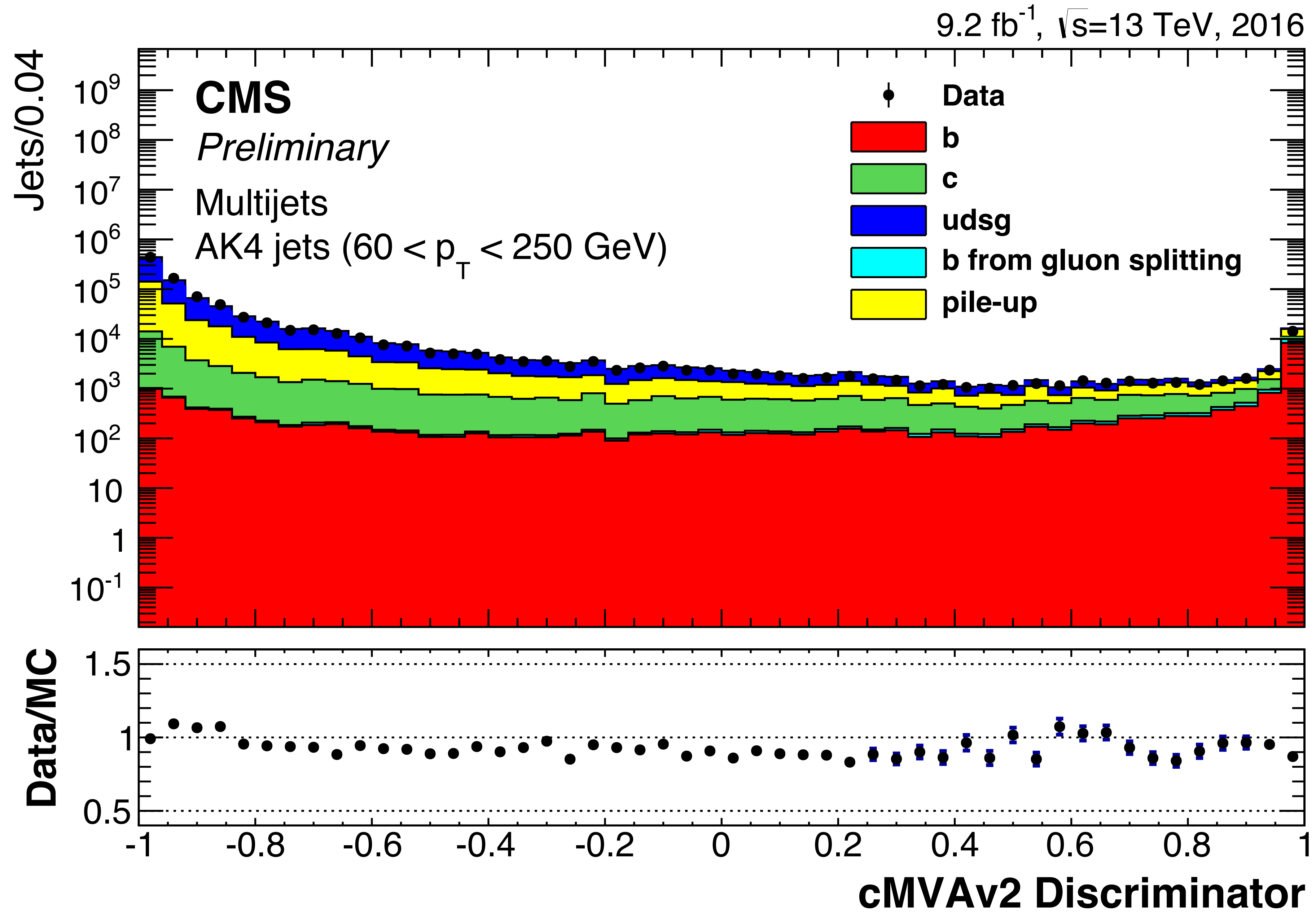
Vertex information

The **Combined Multivariate Algorithm** (cMVAv2) algorithm combines:

- CSVv2 and soft lepton taggers

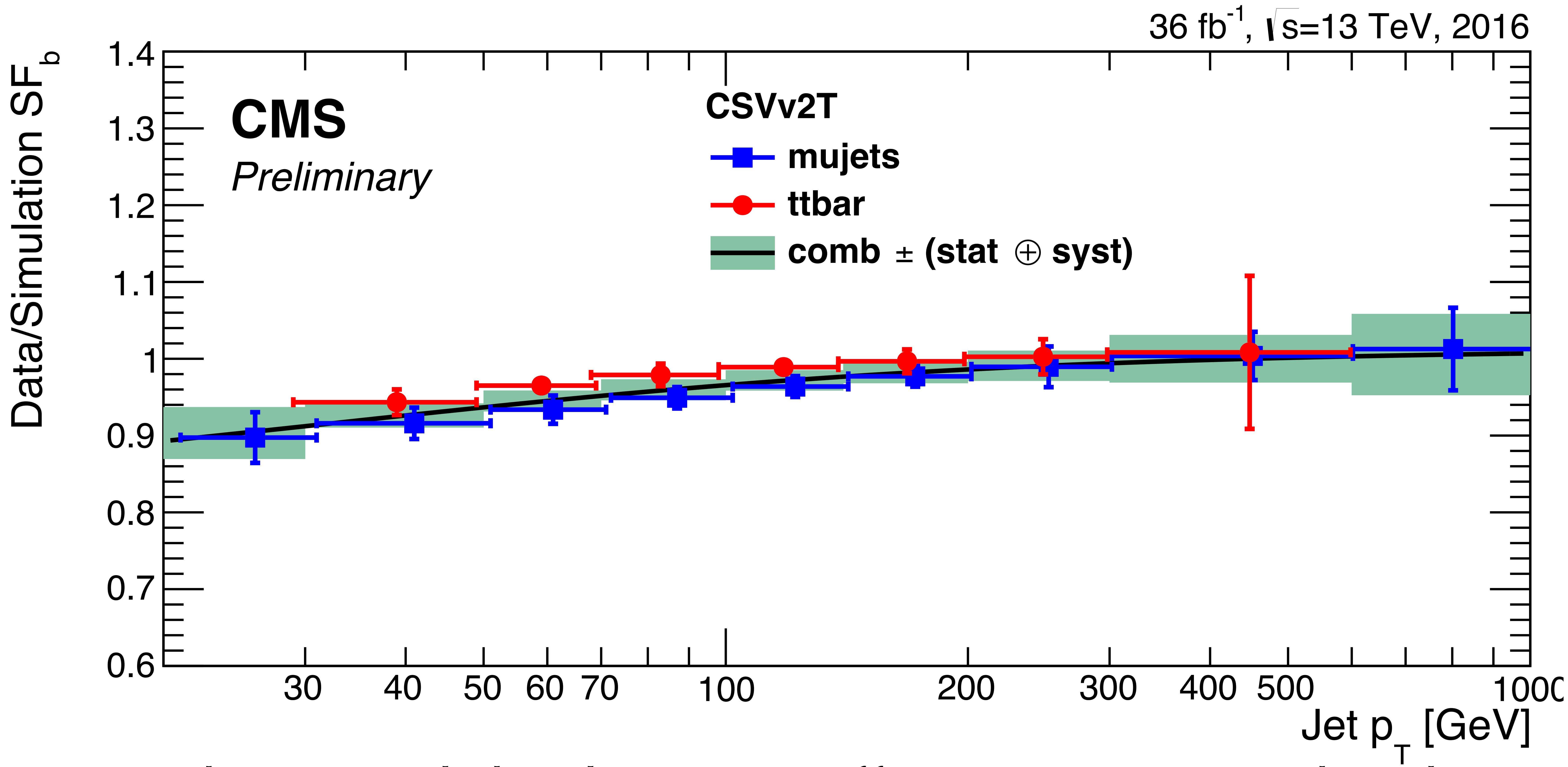
DP-16-042





https://twiki.cern.ch/twiki/pub/CMSPublic/BTV13TeVLHCP2017

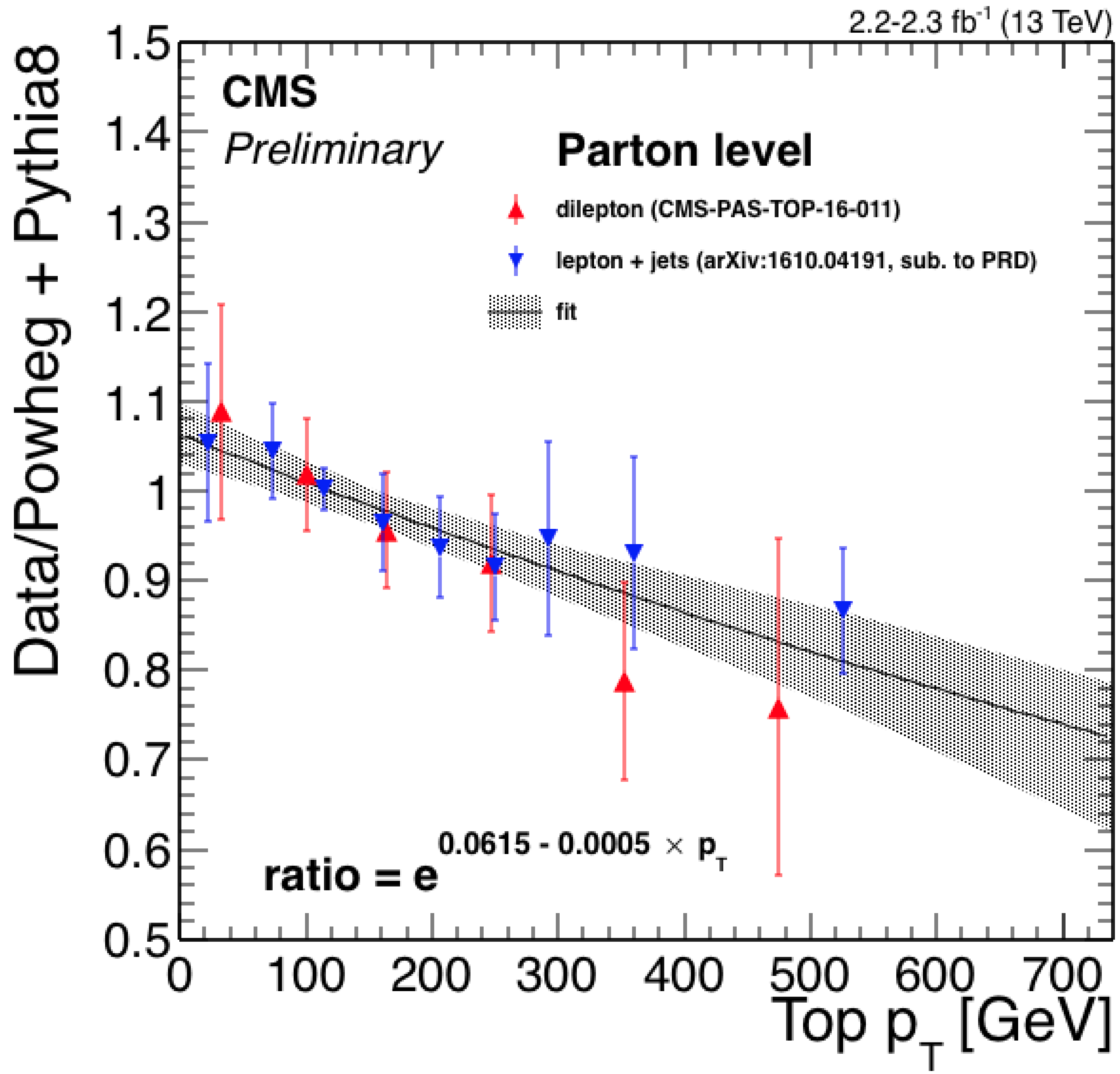
Muon Tagged jets



Good compatibility between efficiency measured with:
- muon tagged jets from multijet events
- b-jets from ttbar

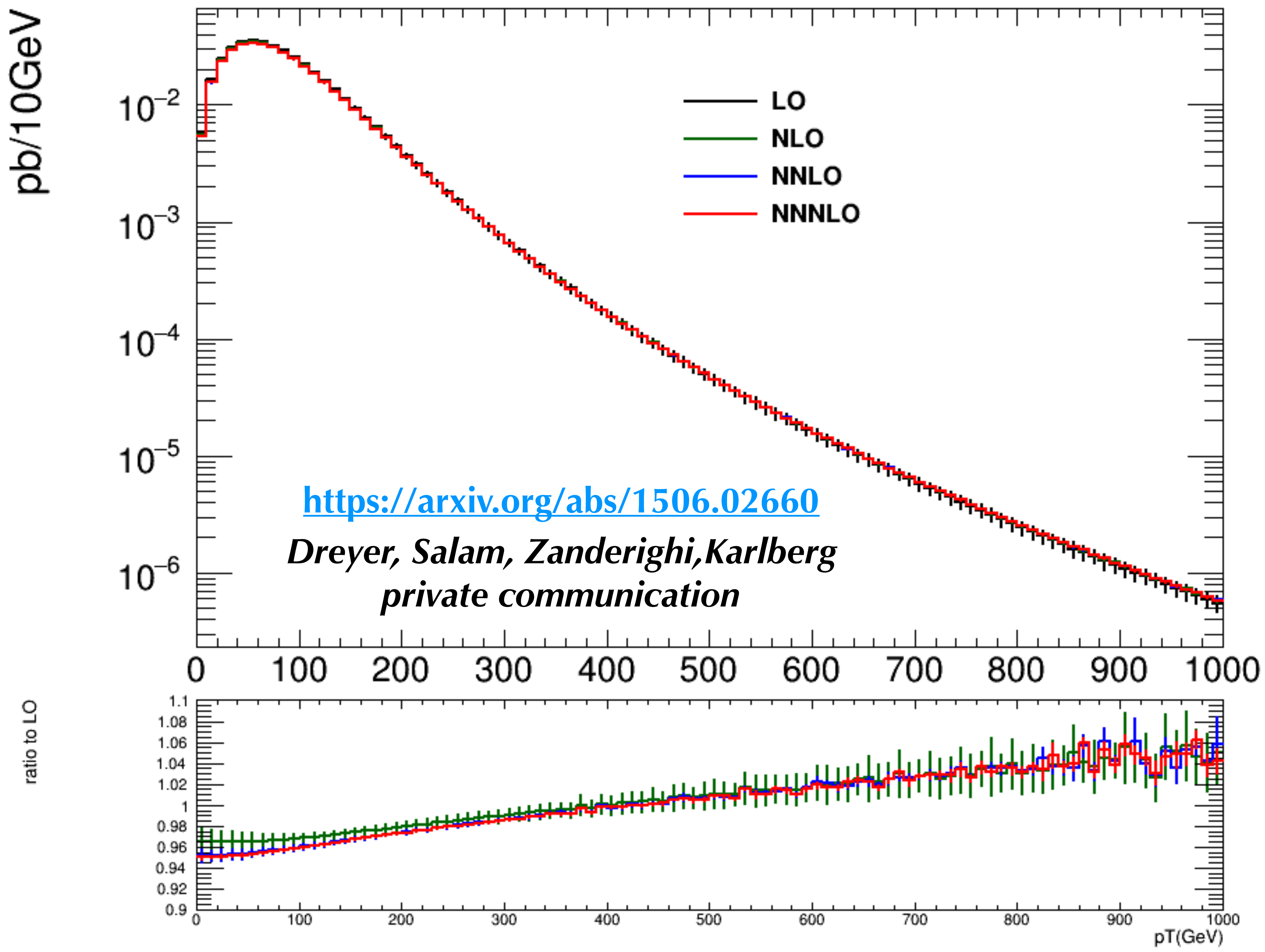
Top p_T reweighting

- Applying 13 TeV top p_T reweighting

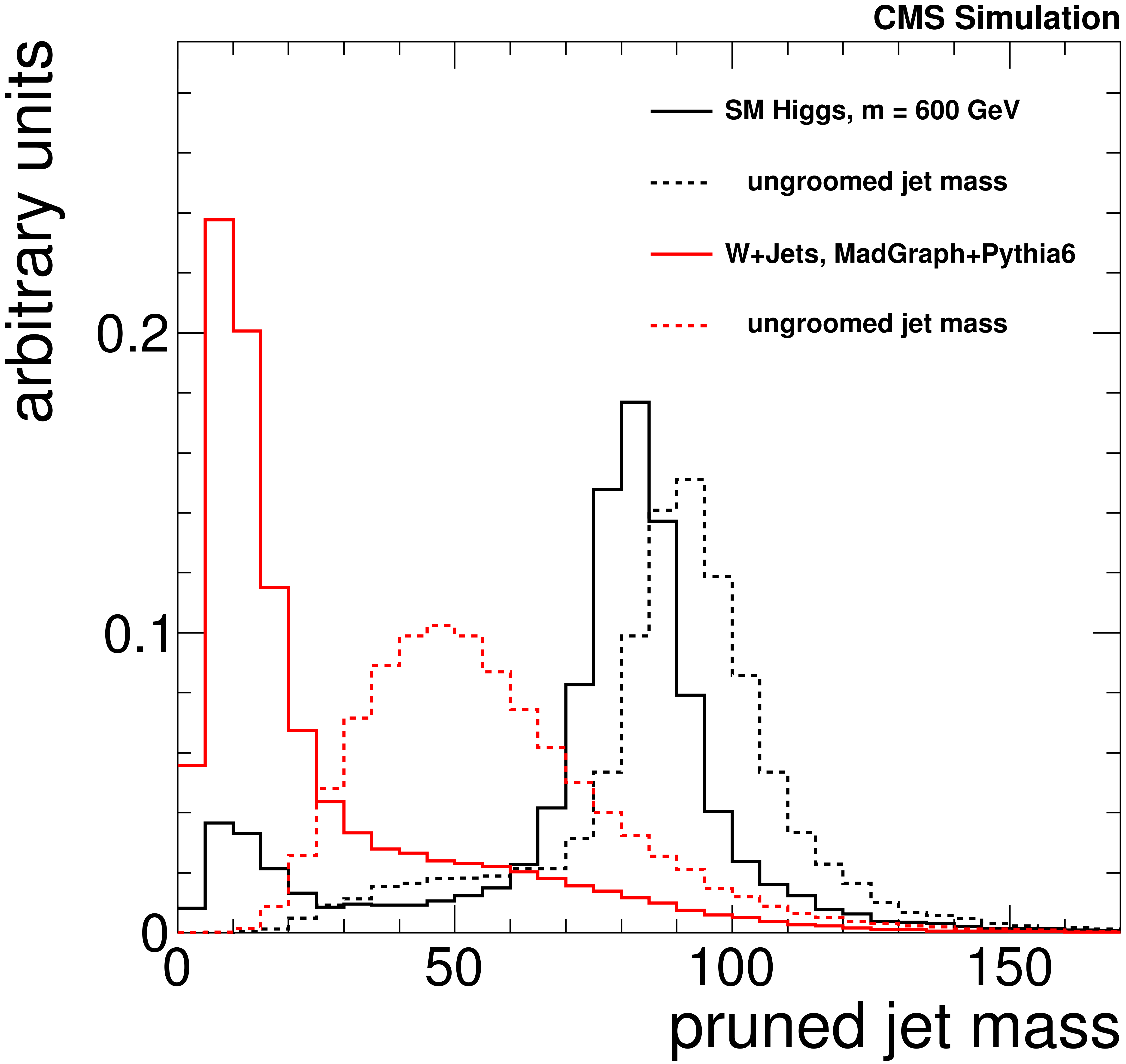


VBF H reweighting

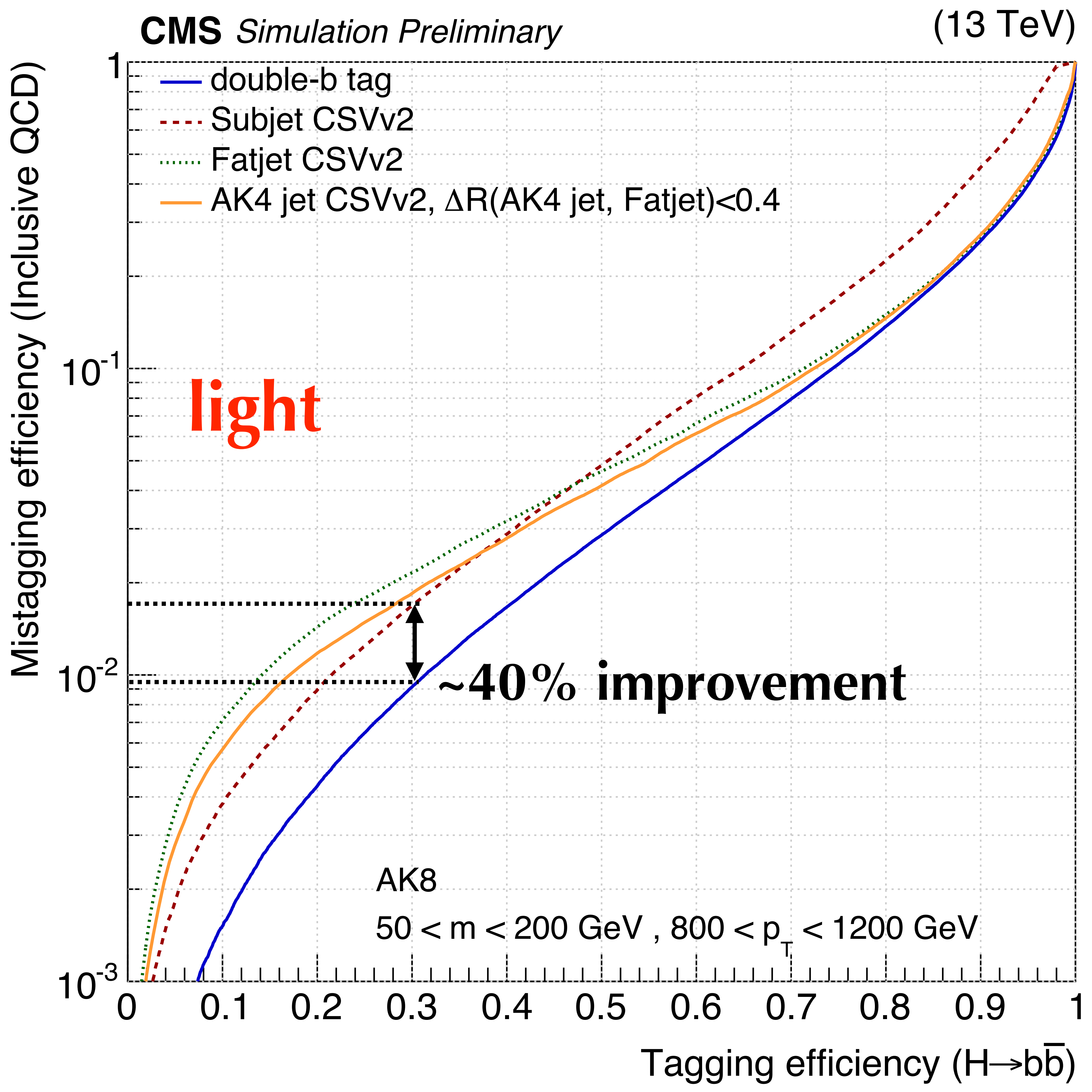
N3LO for VBF have 5-10% effect in our phase space

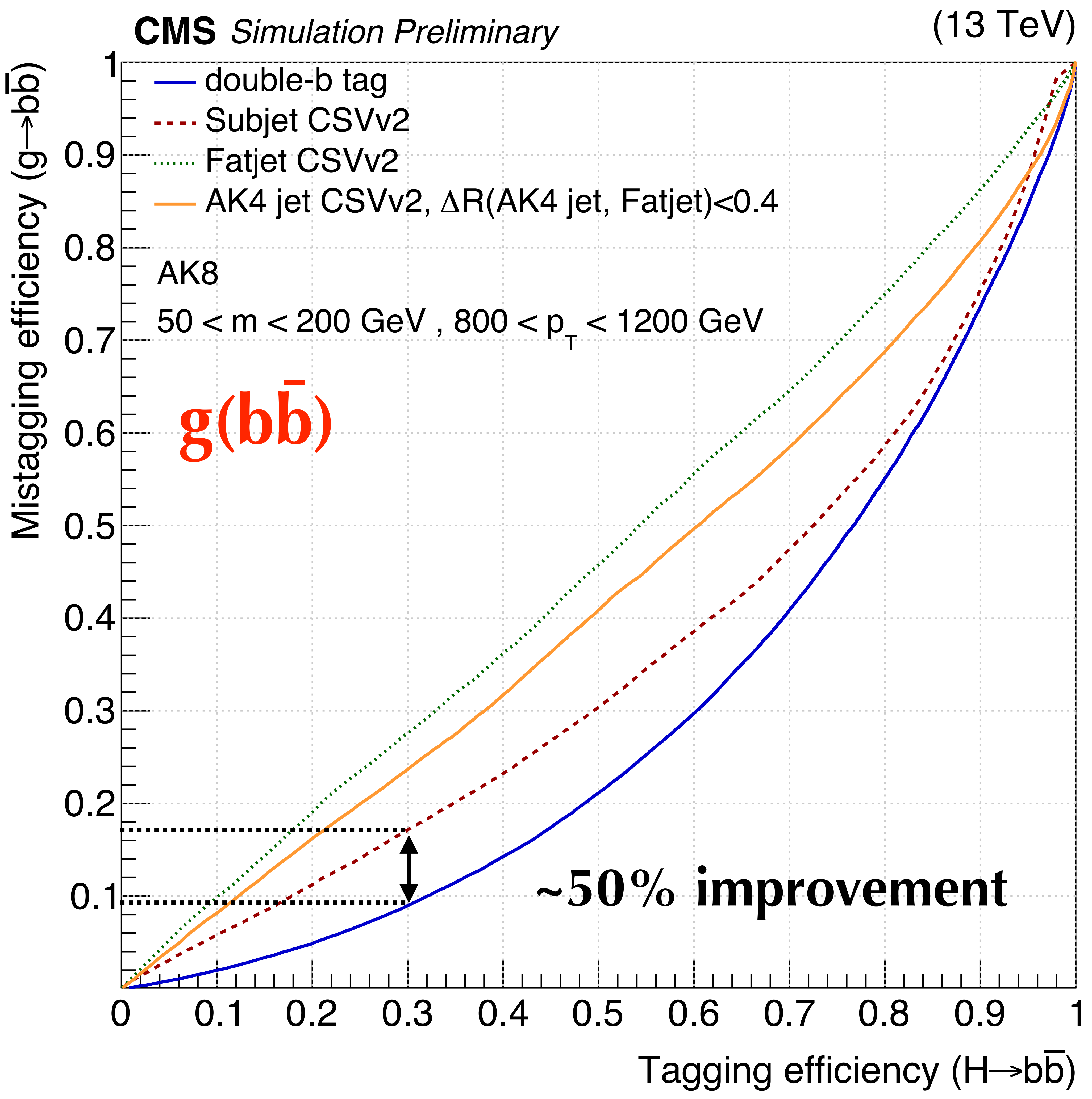


Jet mass grooming



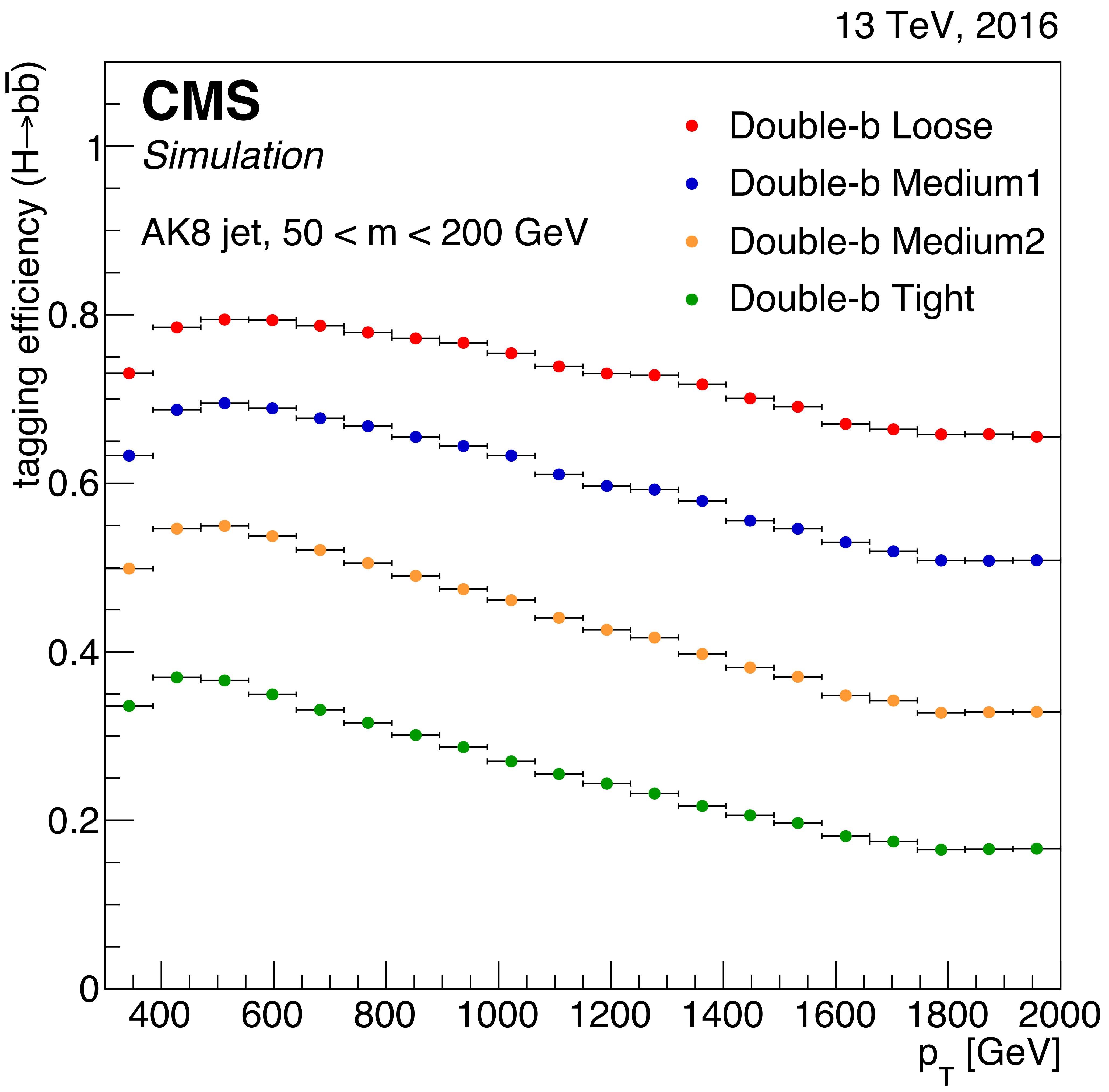
Efficiency vs. Mistag rate

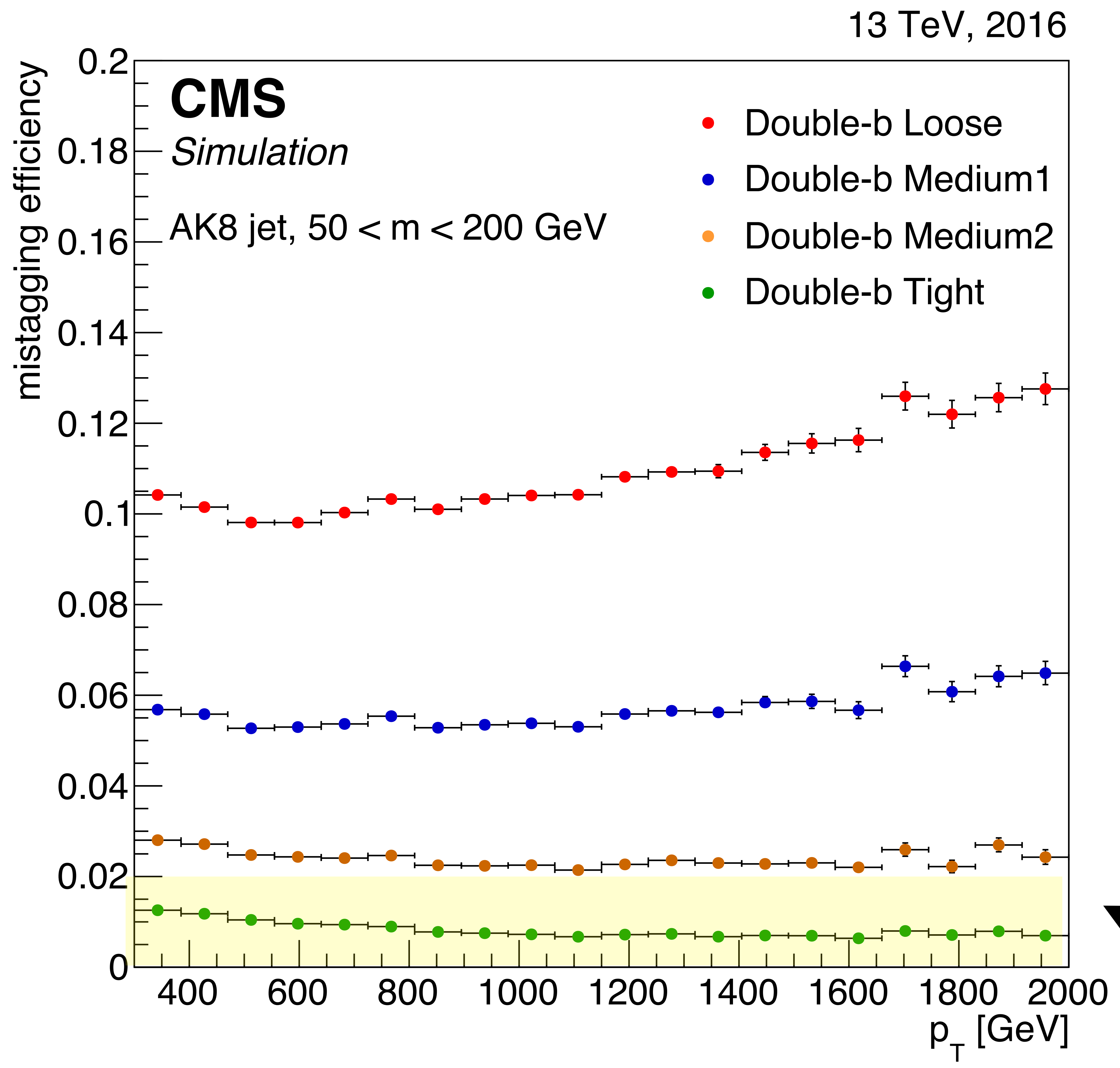




Mistag is reduced by more than 40% at 30% signal efficiency (~ tight working point)

Performance



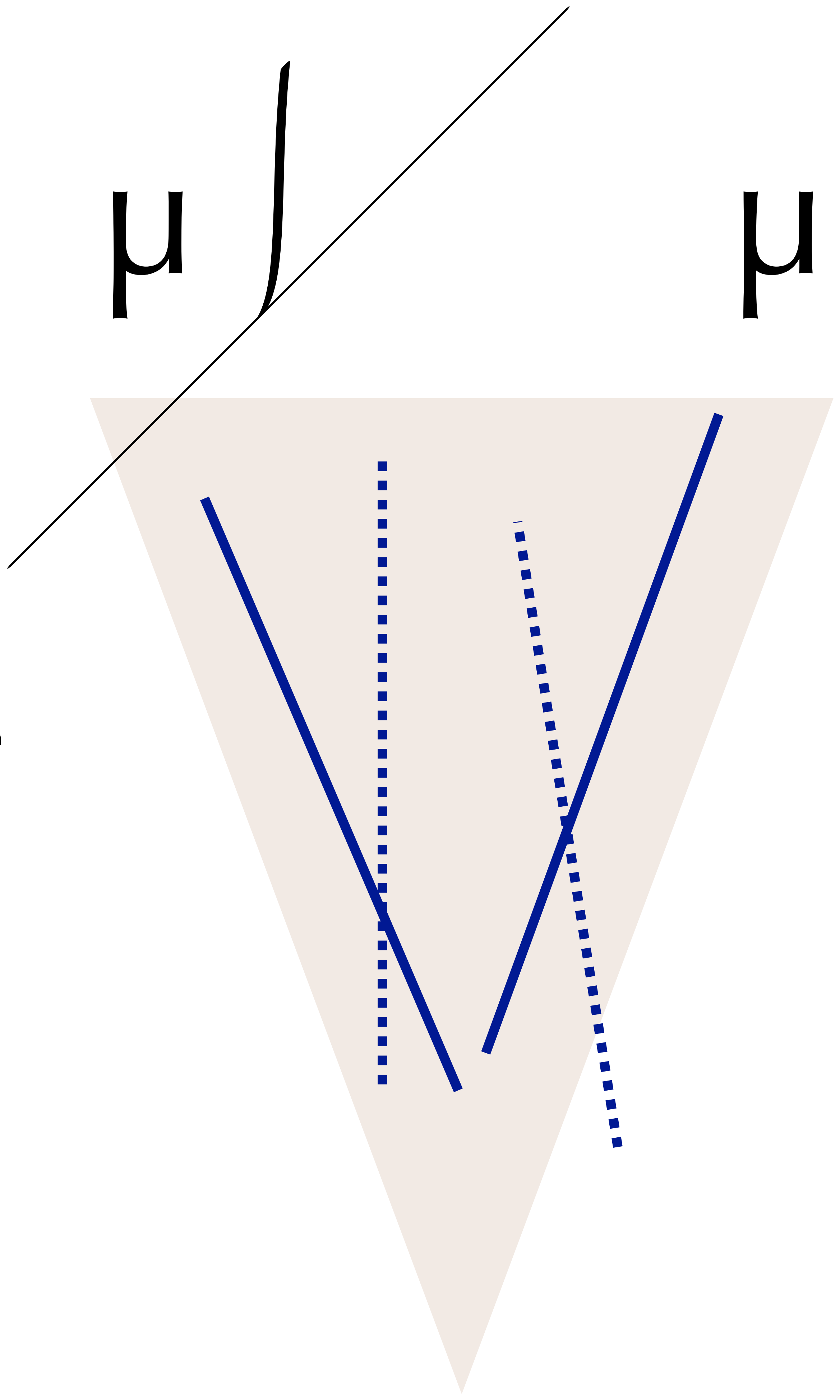


The mistag rate is approximately flat across the p_T range by design

Critical point for searches (background estimate)

Efficiency measurement in data

- Since there is no H/Z($b\bar{b}$) signal (yet!) we use:
  - **g($b\bar{b}$) jets as a proxy** to measure the signal efficiency
- Jet selection has been designed to ensure jets are signal-like
  - High AK8 p_T jet (p_T > 250 GeV)
  - **double-muon** tagged jets (muon with p_T > 7 GeV)
  - **mass cut** (>50 GeV)

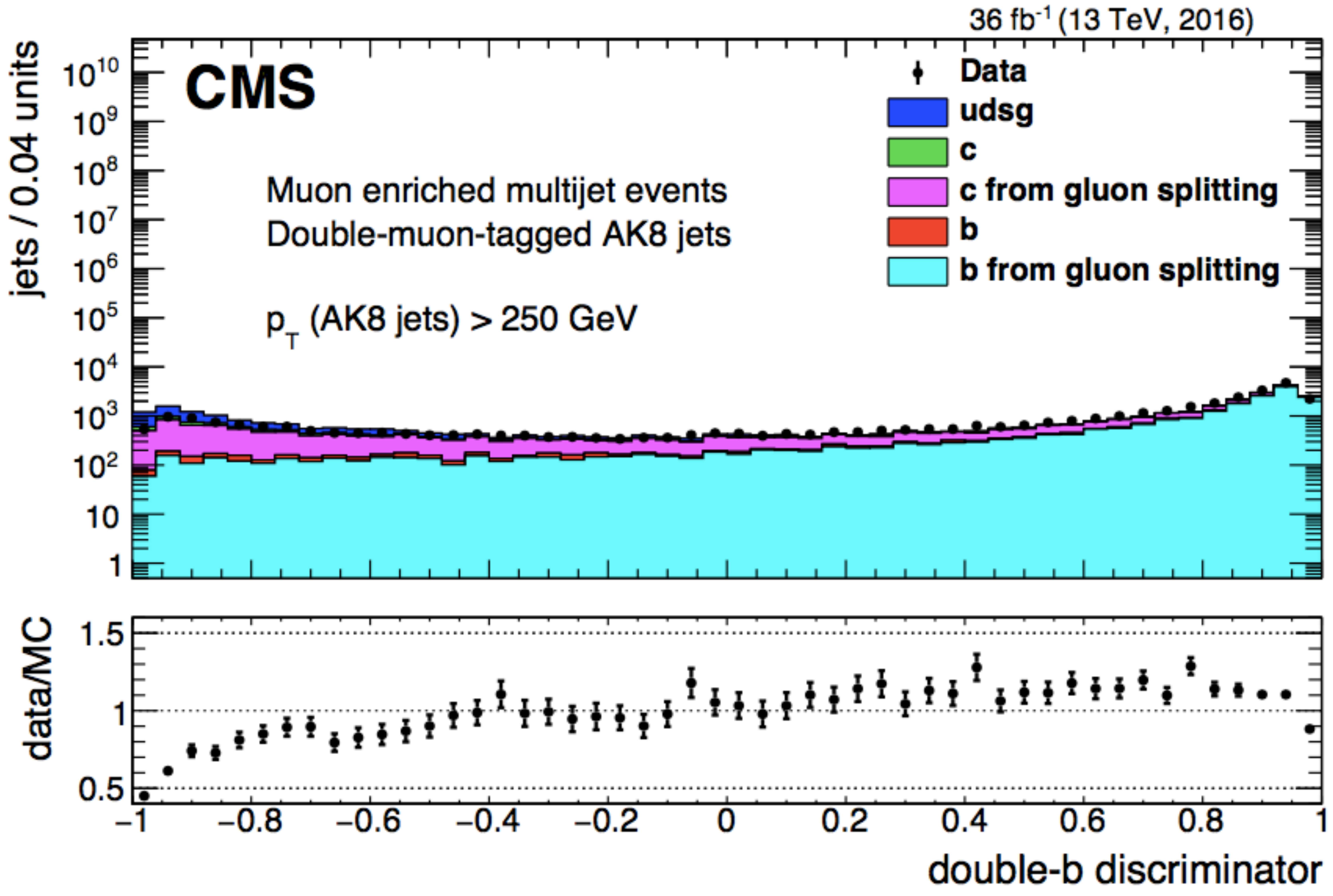


Z($b\bar{b}$) by the end of the talk

Efficiency measurement in data

CMS-PAS-BTV-16-002

Associated **uncertainty** varies from **3 to 5%** depending on the different tagging efficiency

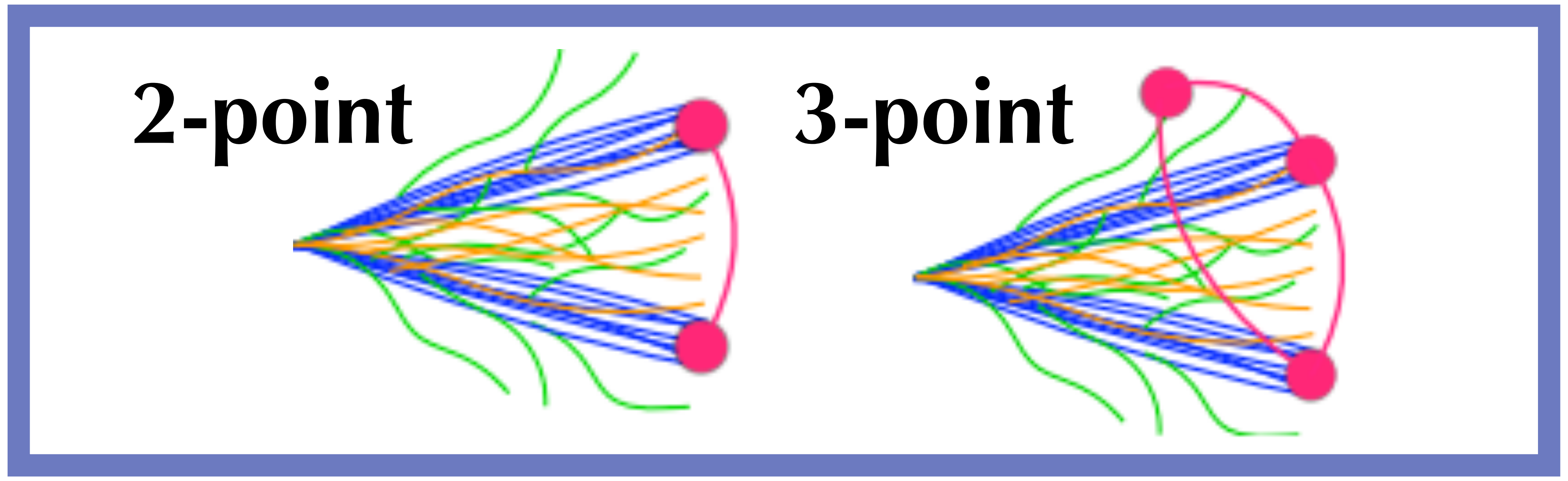


Jet Substructure

ArXiv:1609.07483

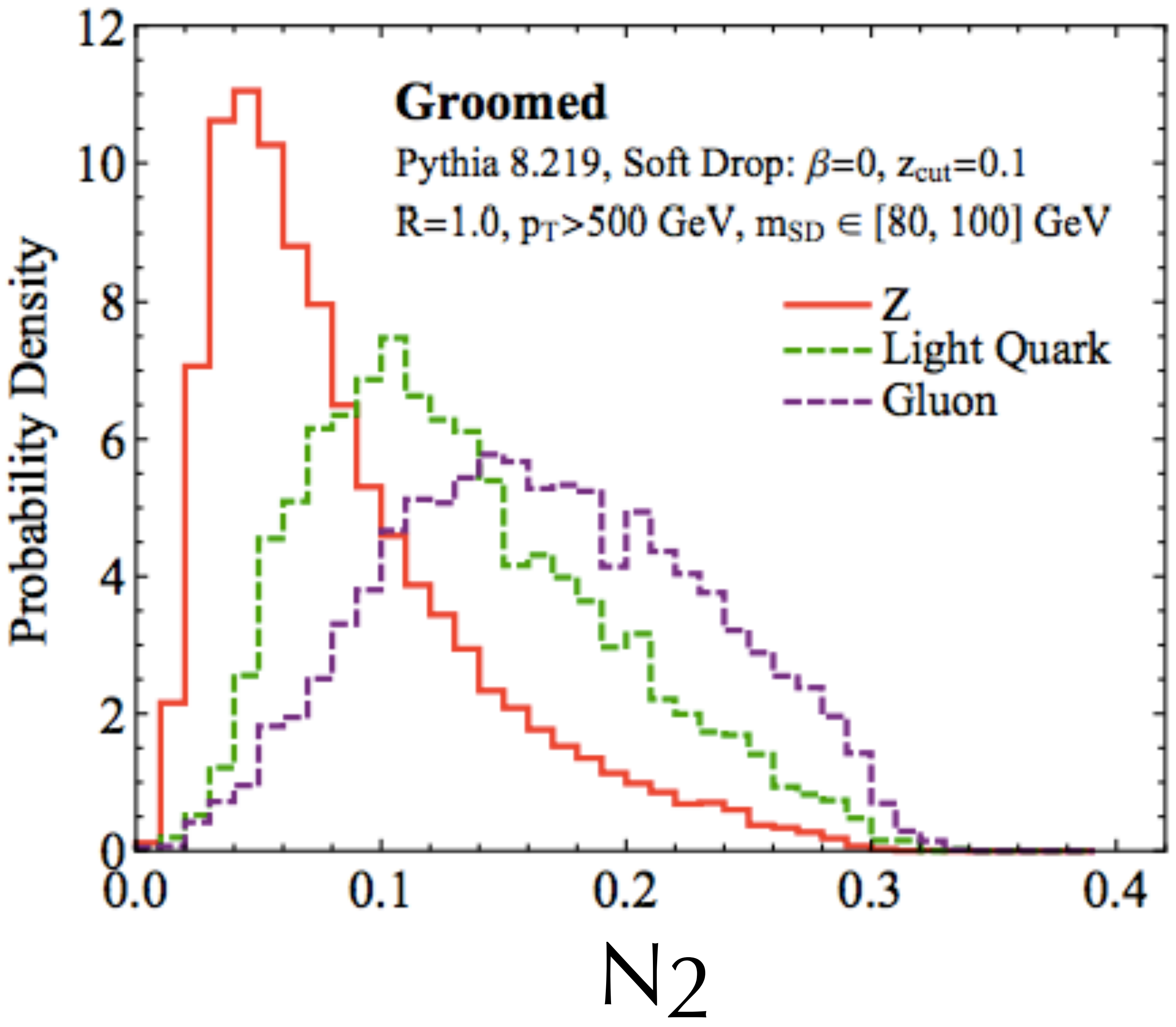
- Measures the degree to which a jet can be considered as composed of N prongs
- **Energy correlation functions** are sensitive to N-point correlations in a jet
  - A 2-pronged jet will have $e_3<e_2$

$$N_2^\beta = \frac{{}_2e_3^\beta}{({}_1e_2^\beta)^2}$$



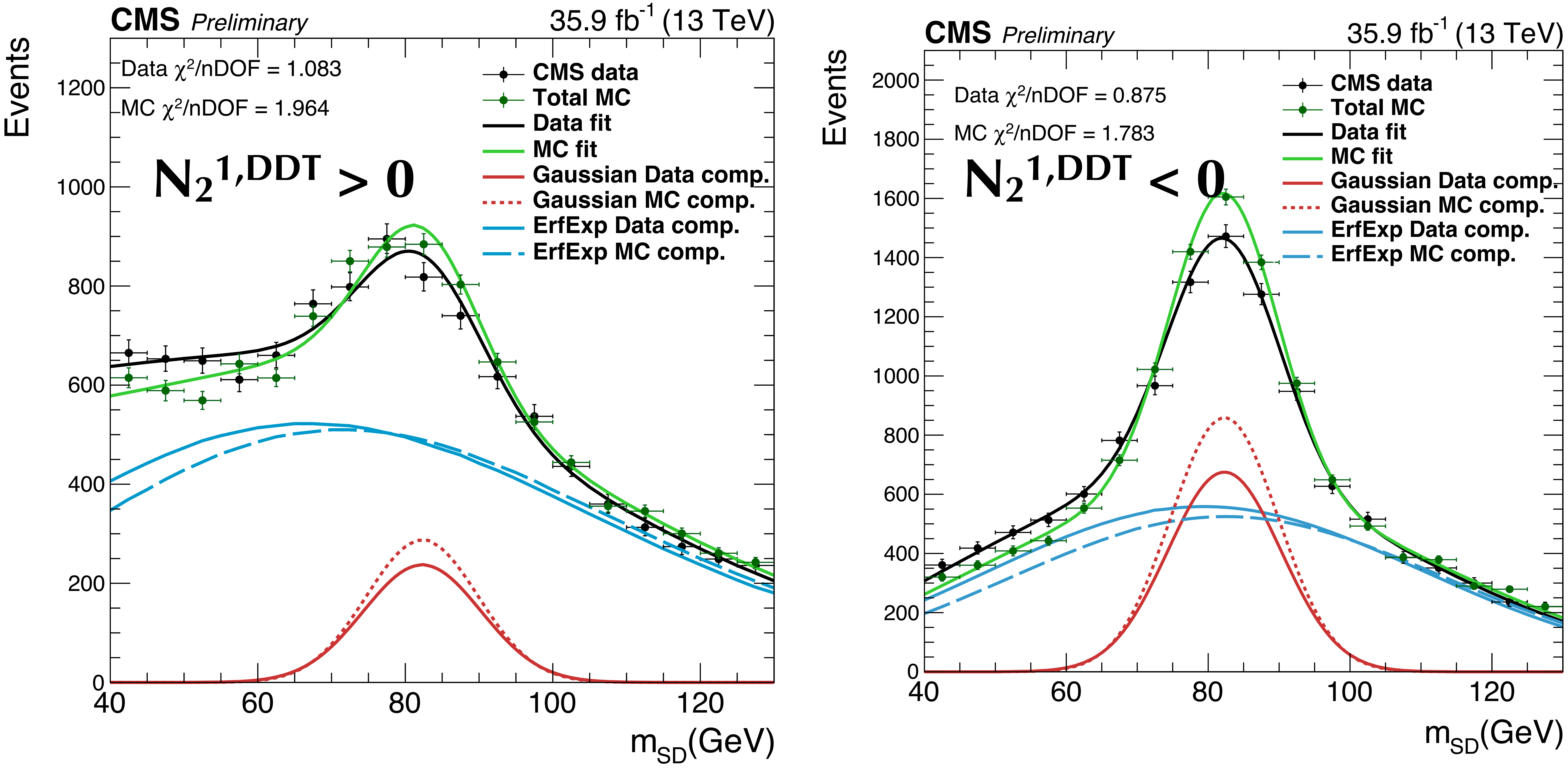
$${}_1e_2^\beta = \sum_{1\leq i<j\leq n_j} z_i z_j \Delta R_{ij}^\beta \qquad z_i = \frac{p_{T_i}}{\sum_{j\in \text{jet}} p_{T_j}} \qquad \beta = 1$$

$${}_2e_3^\beta = \sum_{1\leq i<j<k\leq n_j} z_i z_j z_k \min\left\{\Delta R_{ij}^\beta \Delta R_{ik}^\beta, \Delta R_{ij}^\beta \Delta R_{jk}^\beta, \Delta R_{ik}^\beta \Delta R_{jk}^\beta\right\}$$



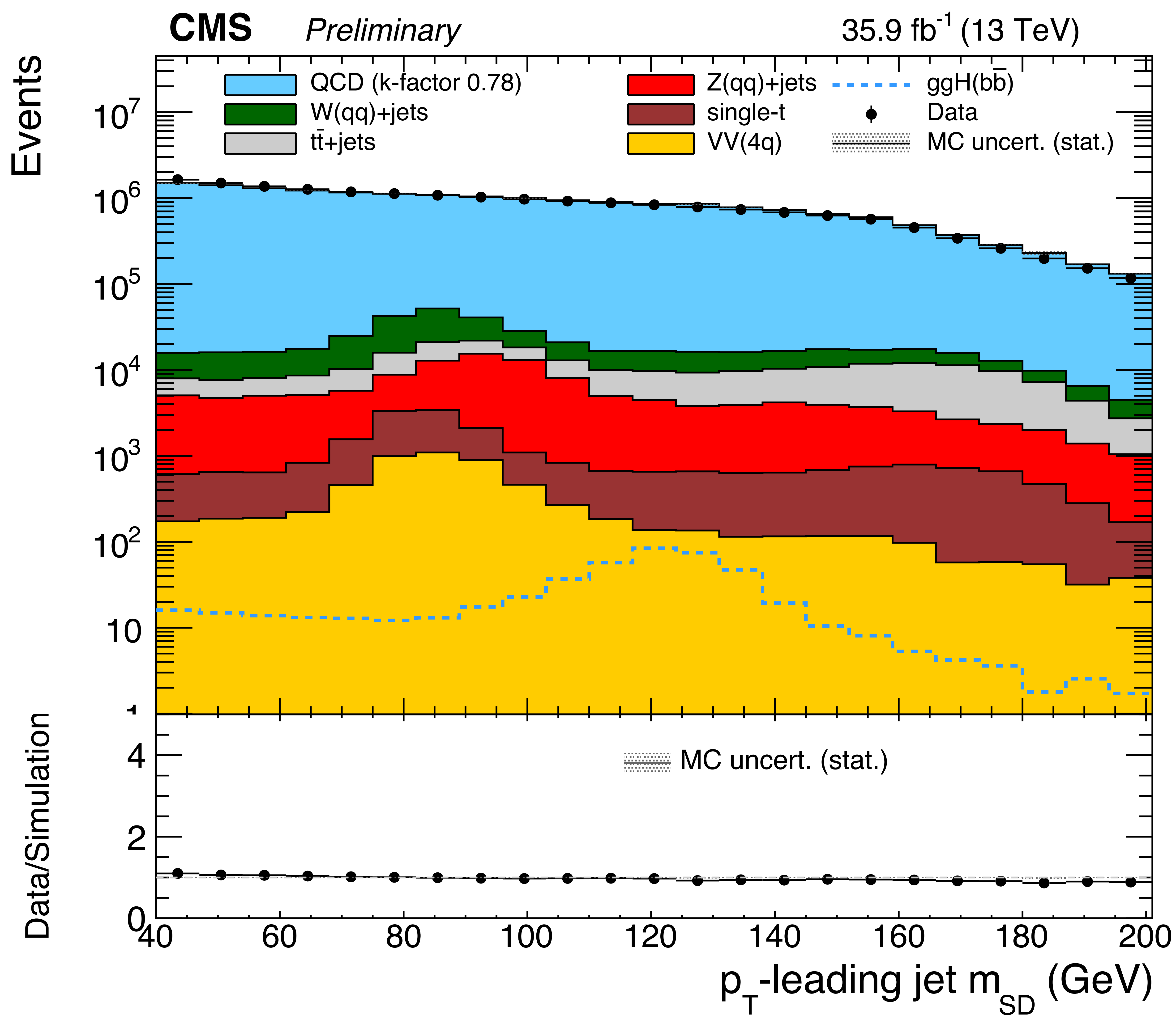
$N_2^{1,DDT}$ efficiency measurement

- Efficiency of the $N_2^{1,DDT}$ is measured in data using merged W jets from tt events



- efficiency SF = $\varepsilon_{Data} / \varepsilon_{MC}$ = 0.993 +/- 0.043
- mass scale SF = m_{Data} / m_{MC} = 1.001 +/- 0.004
- mass resolution SF = $\sigma_{Data} / \sigma_{MC}$ = 1.08 +/- 0.09

Background composition



Backgrounds

- Multijets
- tt+jets, normalization from a dedicated control region

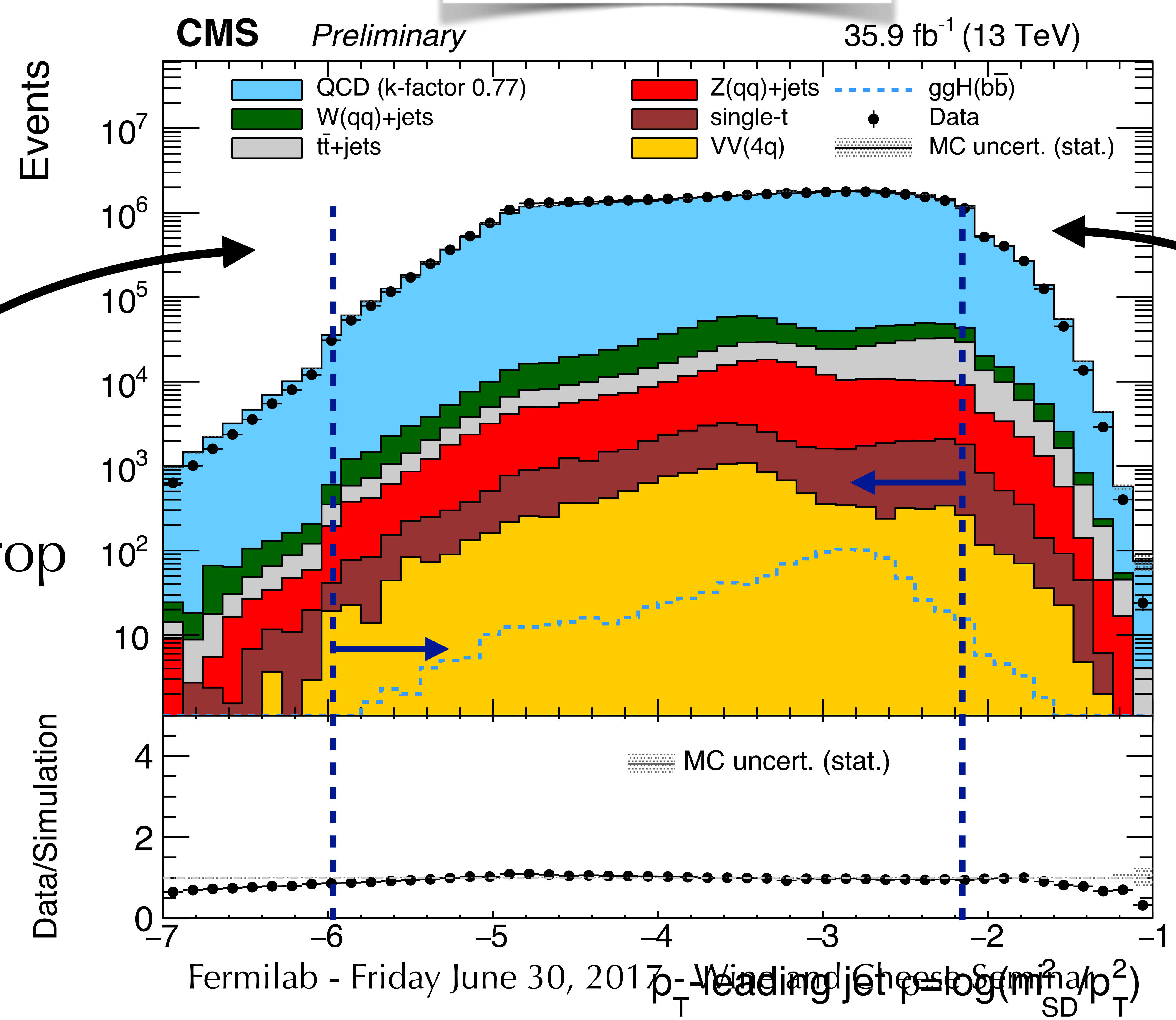
} **data**

- W/Z+jets
- single-t, Diboson

} **MC**

Event Selection

$-6 < \rho < -2.1$



non-perturbative regime of the soft drop mass calculation

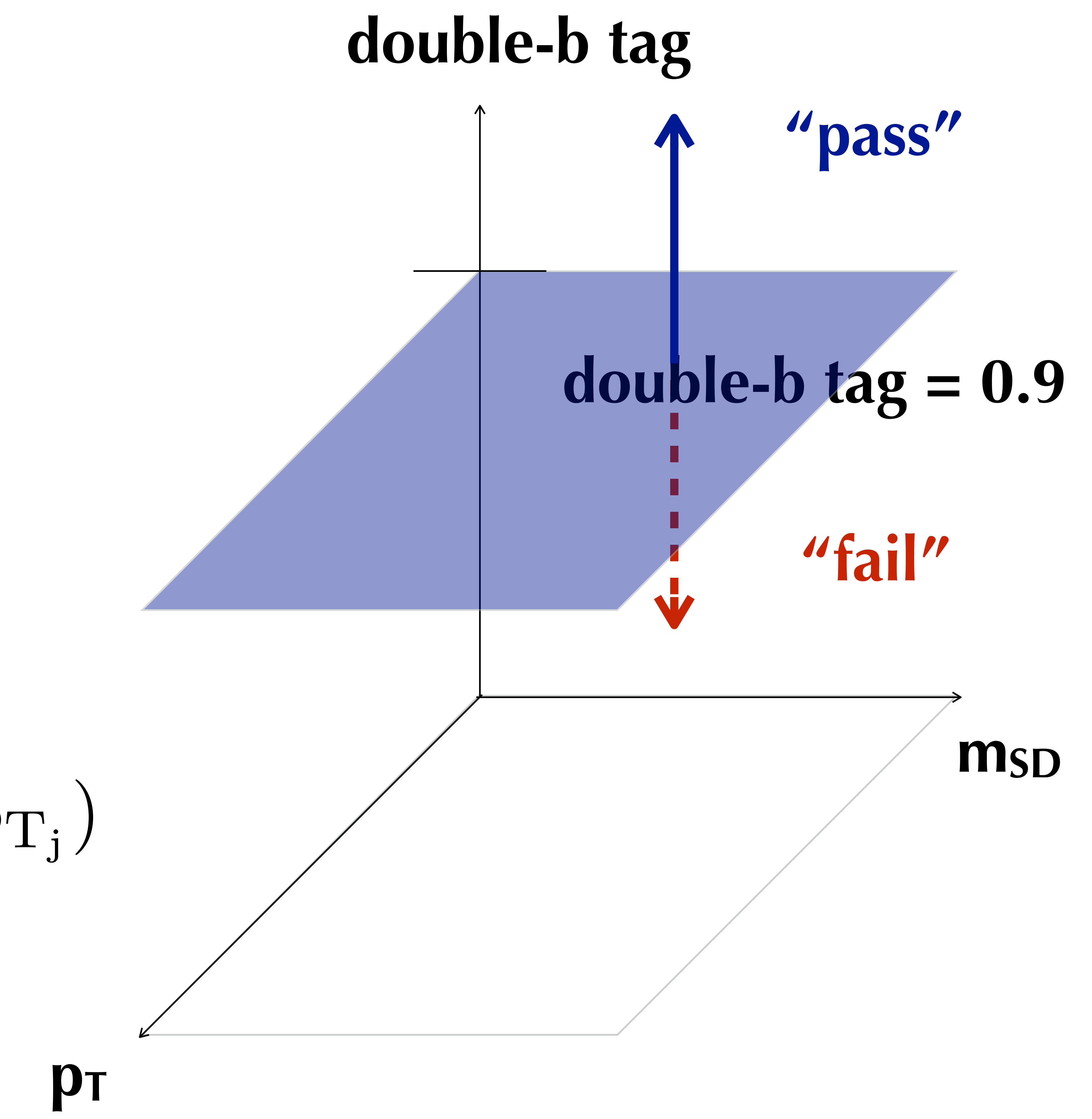
finite cone effects from the AK8 jet clustering

Transfer factor

- If the double-b tagger were completely uncorrelated from jet p_T and m_{SD}, the transfer factor would be flat
  - **Taylor expand** $R_{p/f}$ in ρ and p_T

$$N^{QCD}_{pass}(m_{SD}, p_T) = R_{p/f}(\rho, p_T) \times N^{QCD}_{fail}(m_{SD}, p_T)$$

$$N^{QCD}_{pass}(m_{SD_i}, p_{T_j}) = \boxed{\left(\sum_{k,\ell} a_{kl}\rho^{kl}_{ij} p^{\ell}_{T_j}\right)} \times N^{QCD}_{fail}(m_{SD_i}, p_{T_j})$$



- **23 bins** in $\mathbf{m_{SD}}$ from **40 to 201 GeV** and **6 bins** in $\mathbf{p_T}$ from **450 to 1000 GeV**

Signal extraction

Signal extraction and background estimation performed **simultaneously**

$$\mathcal{L}(\text{data}|\mu,\boldsymbol{\theta}) = \prod_{i,j} \text{Poisson}\left(N^{\text{data}}_{\text{fail},i,j} | N^{\text{QCD}}_{\text{fail},i,j} + N^{t\bar{t}}_{\text{fail},i,j} + N^{V}_{\text{fail},i,j} + \mu\ N^{H(b\bar{b})}_{\text{fail},i,j}\right)$$
$$\times \prod_{i,j} \text{Poisson}\left(N^{\text{data}}_{\text{pass},i,j} | N^{\text{QCD}}_{\text{pass},i,j} + N^{t\bar{t}}_{\text{pass},i,j} + N^{V}_{\text{pass},i,j} + \mu\ N^{H(b\bar{b})}_{\text{pass},i,j}\right)$$

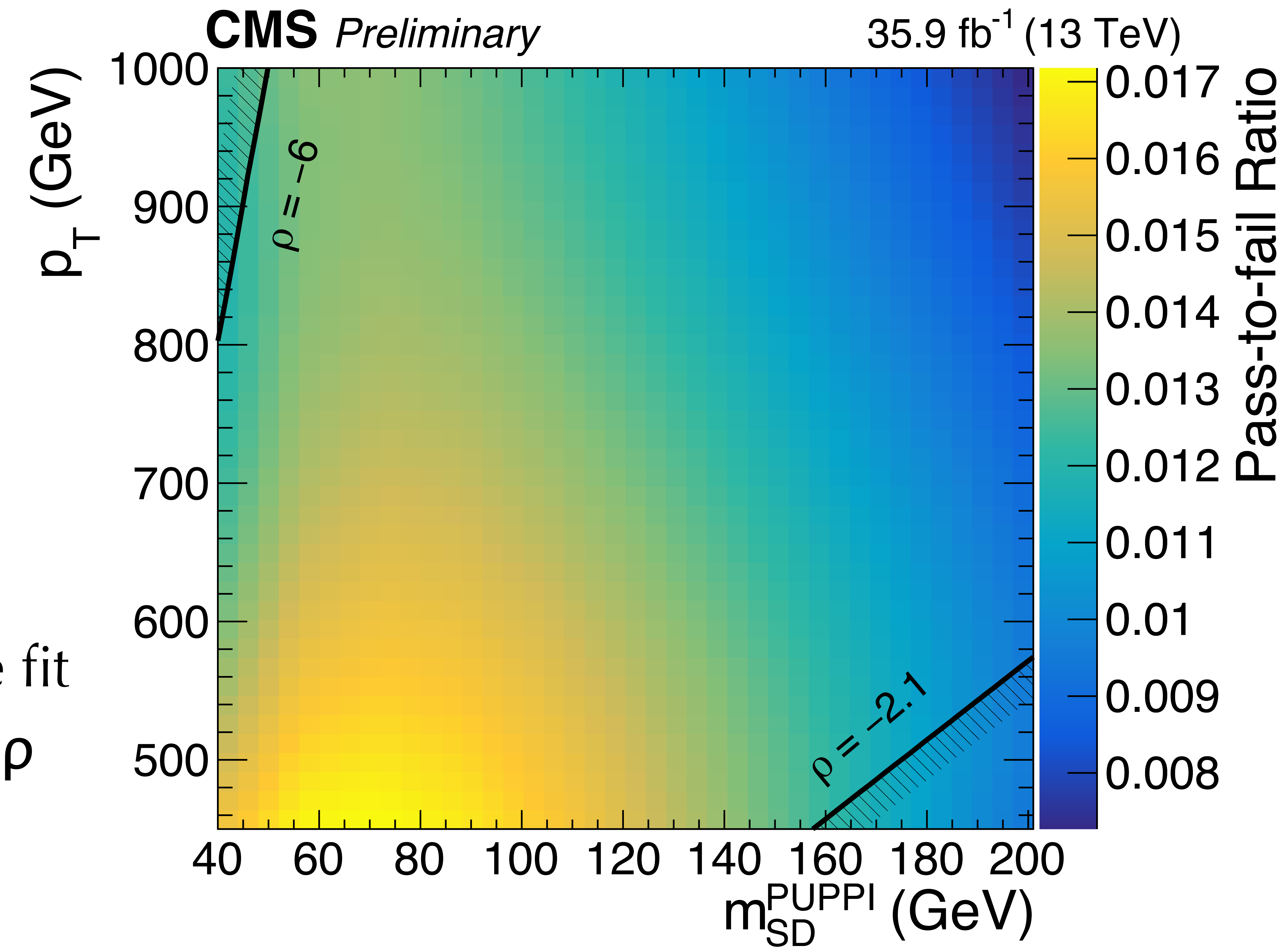
$$\mathrm{N}^{\mathrm{QCD}}_{\mathrm{pass}}(\mathrm{m}_{\mathrm{SD_i}}, \mathrm{p}_{\mathrm{T_j}}) = \boxed{\left(\sum_{\mathrm{k},\ell} \mathrm{a}_{\mathrm{kl}} \rho^{\mathrm{kl}}_{\mathrm{ij}} \mathrm{p}^{\ell}_{\mathrm{T_j}}\right)} \times \mathrm{N}^{\mathrm{QCD}}_{\mathrm{fail}}(\mathrm{m}_{\mathrm{SD_i}}, \mathrm{p}_{\mathrm{T_j}})$$

- The coefficients a_{kl} are determined from the fit
- Based on F-test, a 2nd order polynomial in ρ and 1st order in p_T is used.

Signal extraction

Signal extraction and background estimation performed **simultaneously**

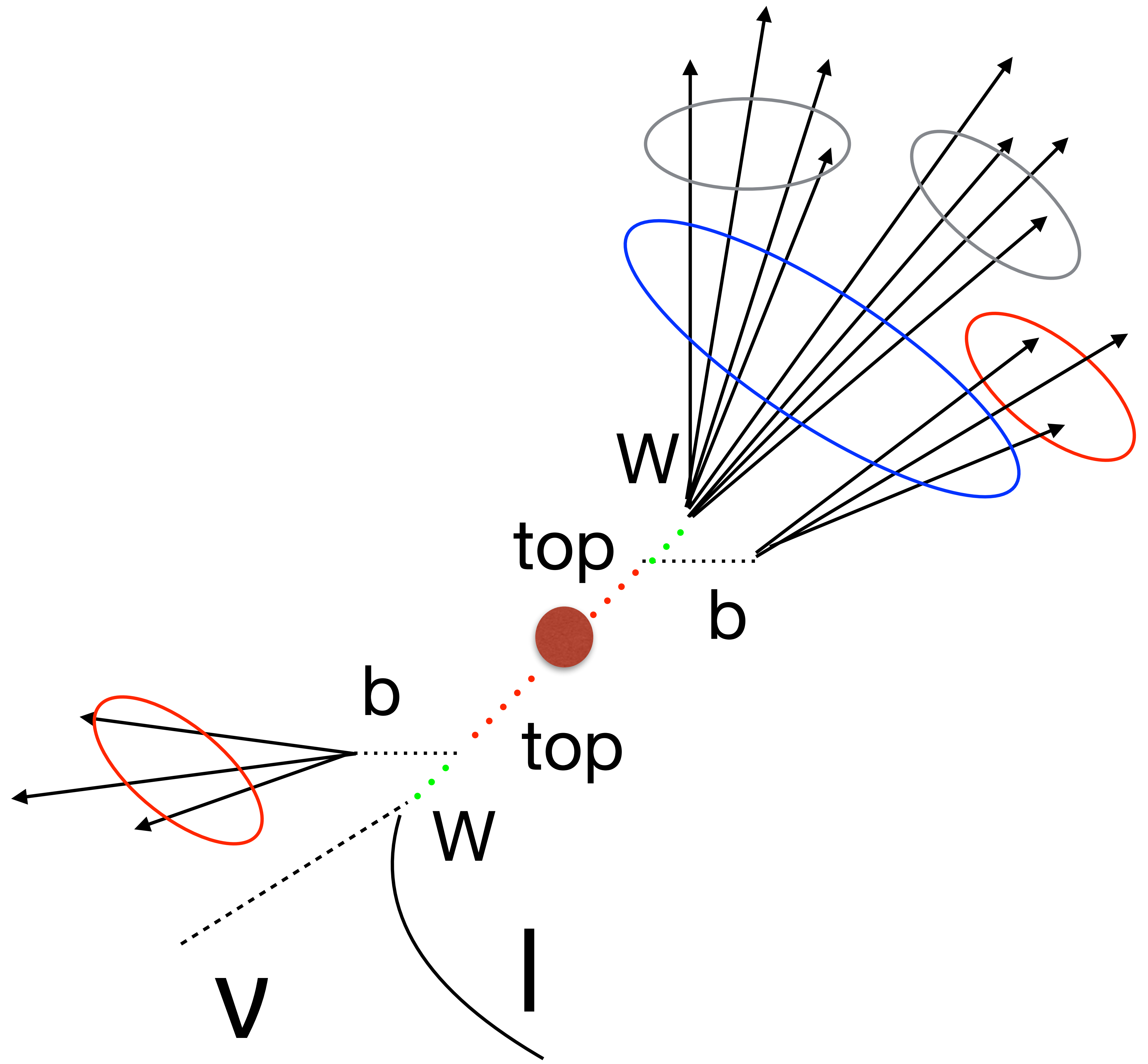
$$N^{QCD}_{pass}(m_{SD_i}, p_{T_j}) = \boxed{\left(\sum_{k,\ell} a_{kl}\rho^{kl}_{ij} p^{\ell}_{T_j}\right)} \times N^{QCD}_{fail}(m_{SD_i}, p_{T_j})$$



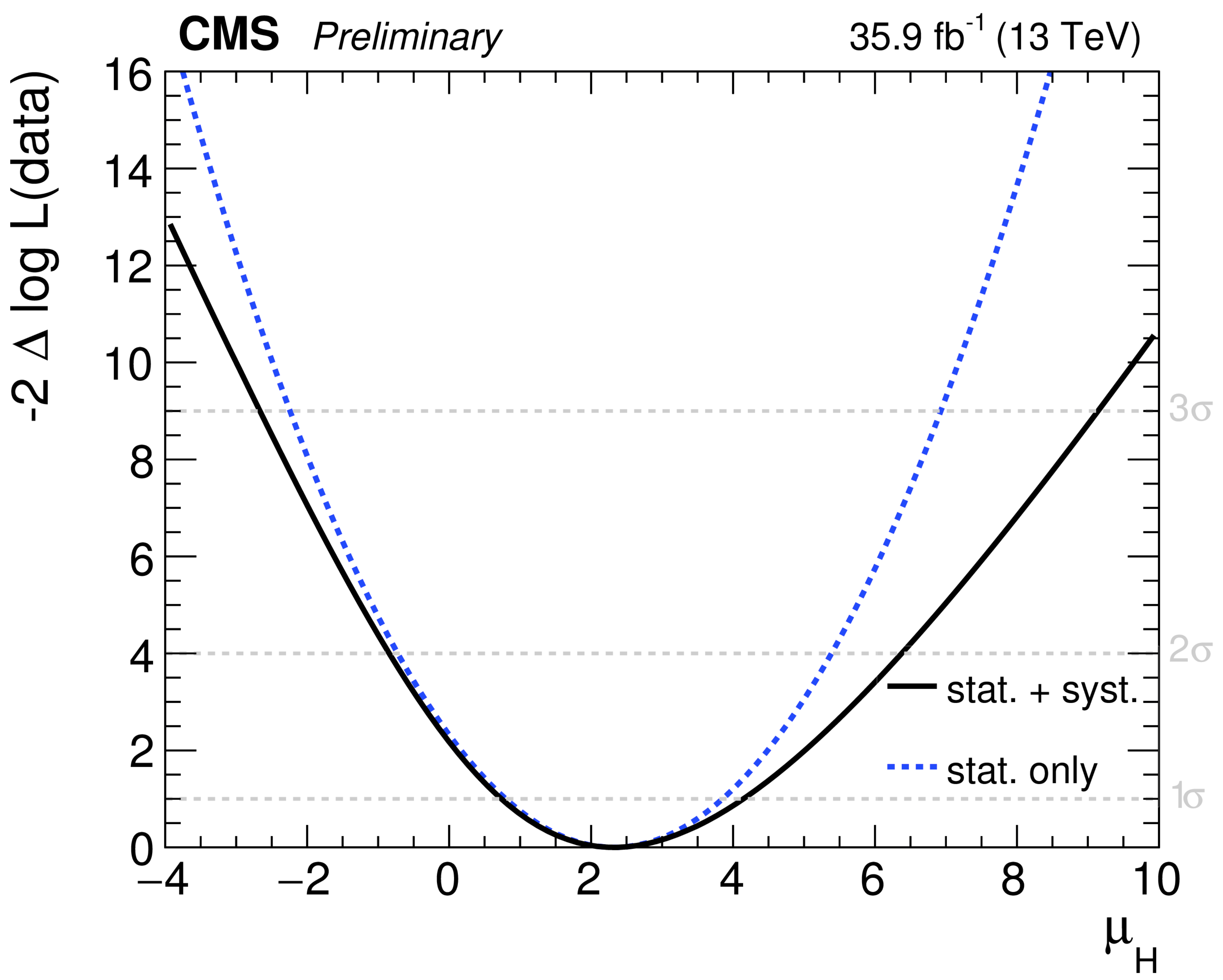
- The coefficients a_{kl} are determined from the fit
- Based on F-test, a 2nd order polynomial in ρ and 1st order in p_T is used.

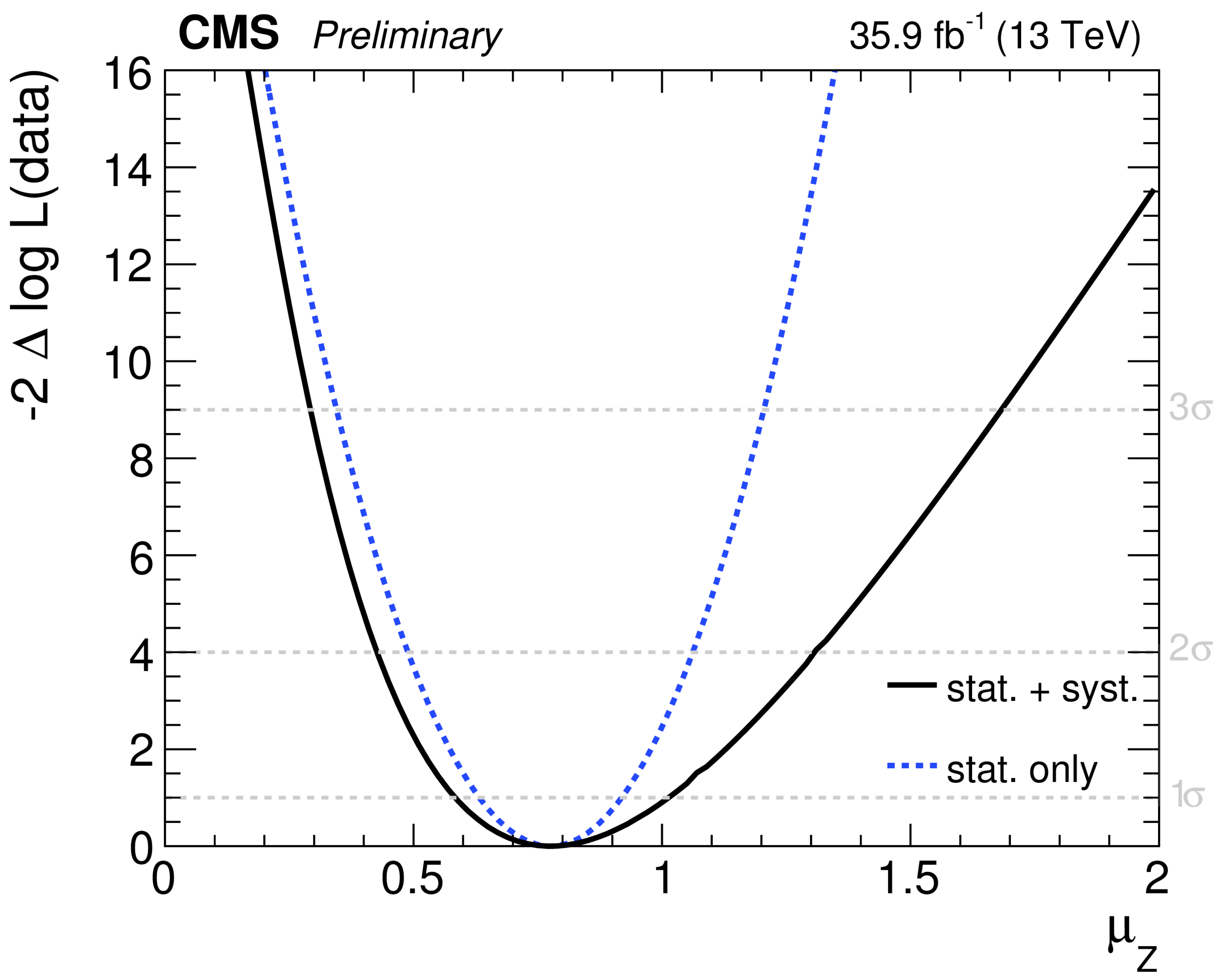
tt enriched control region

- Loose muon with $p_T > 55$ GeV, $|\eta| < 2.1$ in opposite hemisphere:
  - $|\varphi(\mu) - \varphi(\text{AK8 jet})| > 2\pi/3$
- One AK4 PUPPI jet with medium CSVv2 b-tag, with $p_T > 50$ GeV, $|\eta| < 2.5$, and
  - $\Delta R(\text{AK4 b-tag, AK8 jet}) > 0.8$
- Lepton: Veto the presence of identified loose electrons and loose hadronic taus
- Two-prong AK8 PUPPI jet ($p_T > 400$ GeV, $m_{SD} > 40$ GeV and $|\eta| < 2.4$)
  - $N_2^{1,DDT} < 0$



1D LL Scans

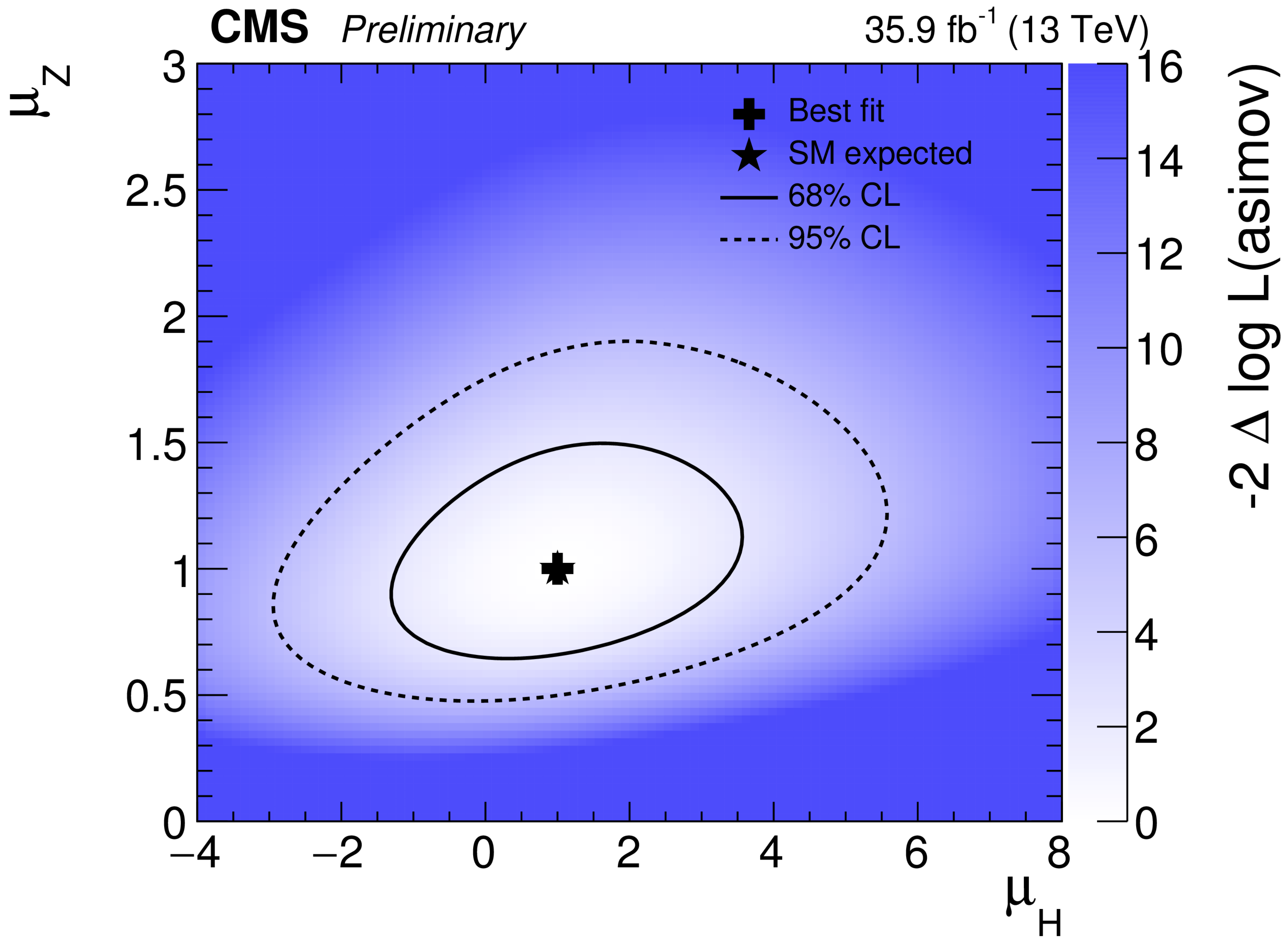




$$\mu_Z = 0.78\ {-0.14}/{+0.14}\ \text{(stat.)}\ {-0.13}/{+0.19}\ \text{(syst.)}$$

$$\mu = 2.3\ {-1.5}/{+1.5}\ \text{(stat.)}\ \ {-0.4}/{+1.0}\ \text{(syst.)}$$

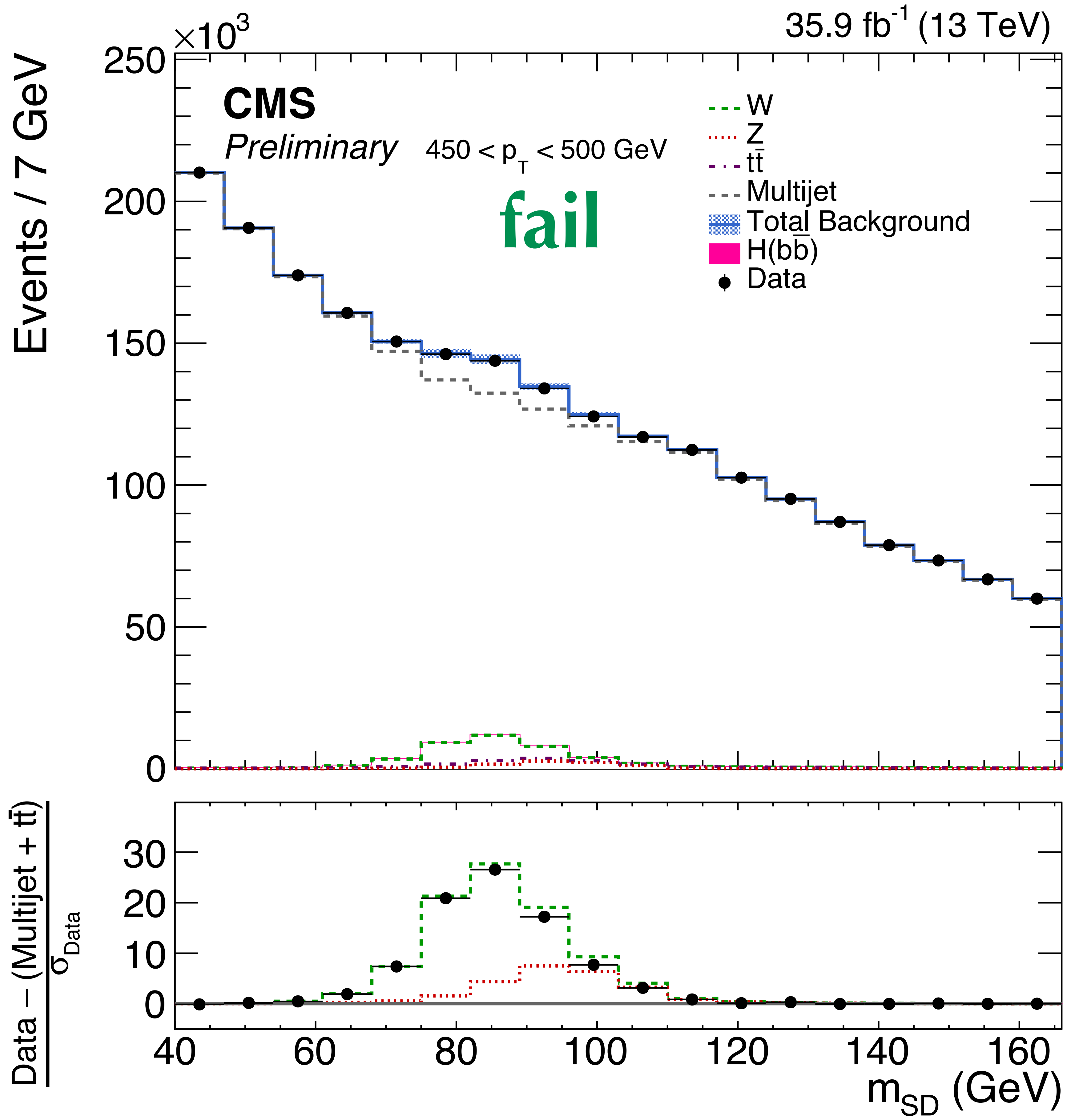
2D LL Scan (Asimov)

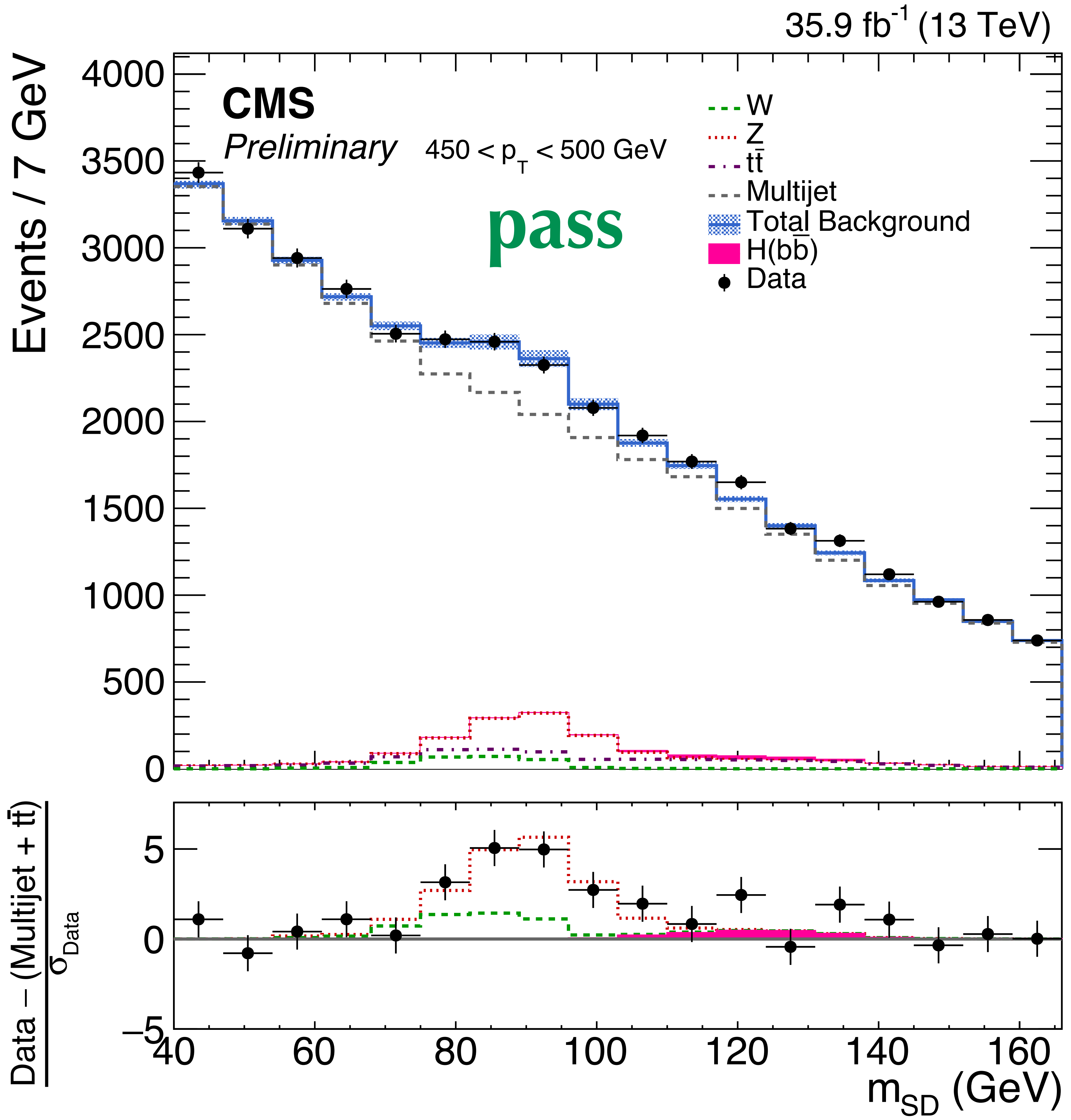


Simultaneous fit of the Z and H signals

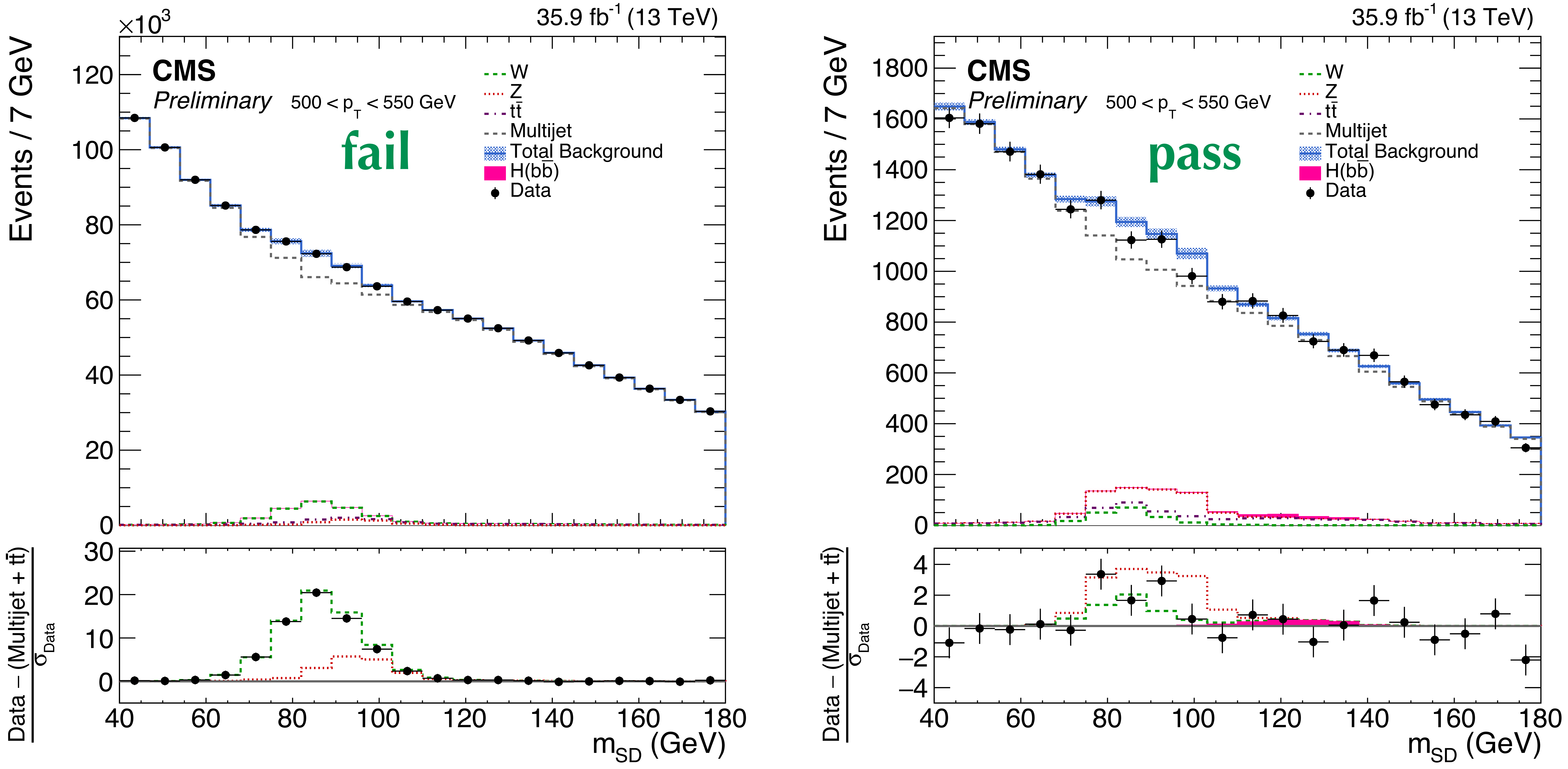
	H	H no p_T corrections	Z
Observed best fit	$\mu_H = 2.3^{+1.8}_{-1.6}$	$\mu'_H = 3.2^{+2.2}_{-2.0}$	$\mu_Z = 0.78^{+0.23}_{-0.19}$
Expected significance	0.7σ ($\mu_H = 1$)	0.5σ ($\mu'_H = 1$)	5.8σ ($\mu_Z = 1$)
Observed significance	1.5σ	1.6σ	5.1σ

Results

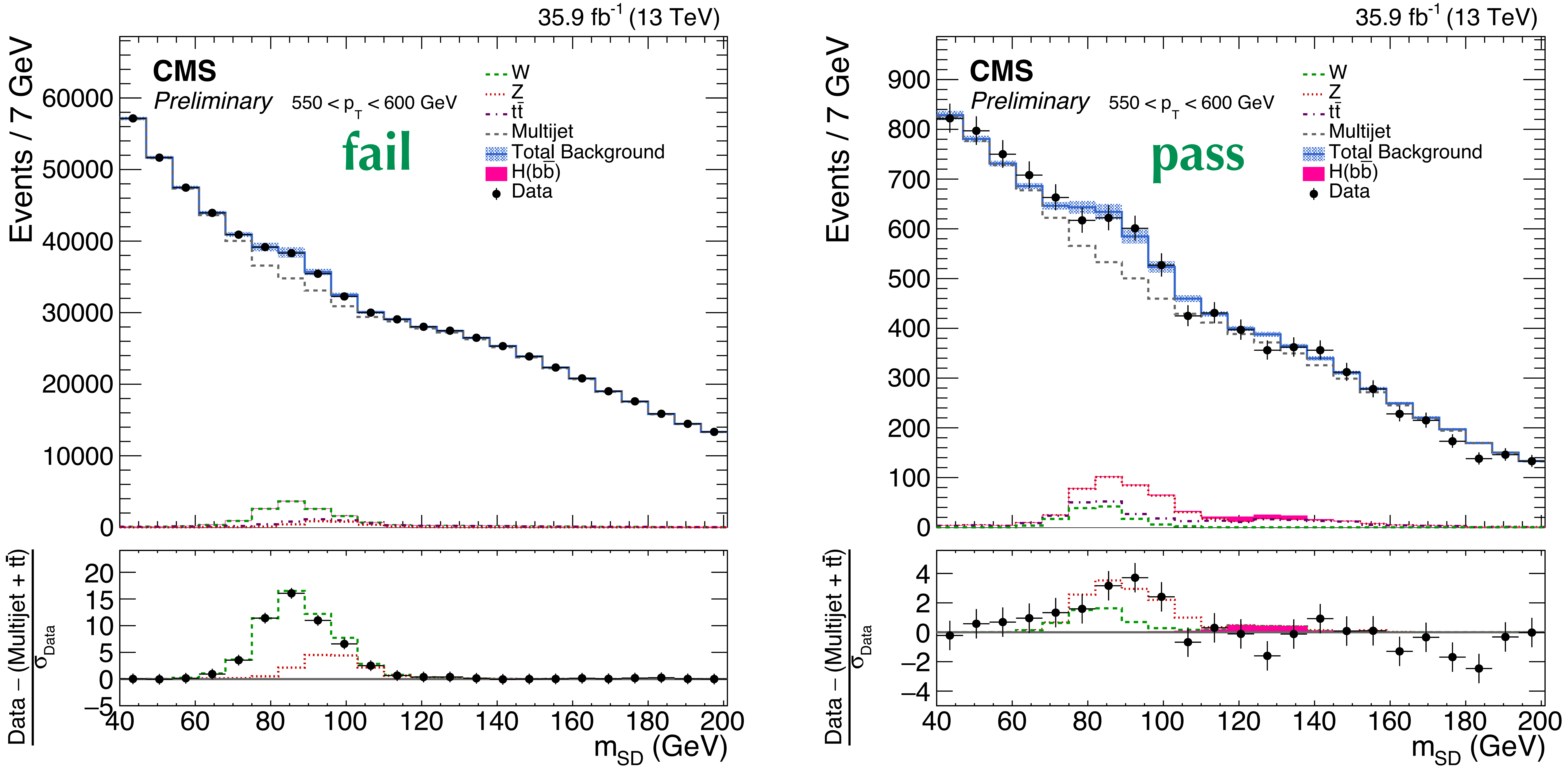

35.9 fb^{-1} (13 TeV)
CMS
Preliminary
450 < p_T < 500 GeV
fail
W
Z
$t\bar{t}$
Multijet
Total Background
H($b\bar{b}$)
Data
Events / 7 GeV
Data – (Multijet + $t\bar{t}$) / σ_{Data}
m_{SD} (GeV)


35.9 fb^{-1} (13 TeV)
CMS
Preliminary
450 < p_T < 500 GeV
pass
W
Z
$t\bar{t}$
Multijet
Total Background
H($b\bar{b}$)
Data
Events / 7 GeV
Data – (Multijet + $t\bar{t}$) / σ_{Data}
m_{SD} (GeV)

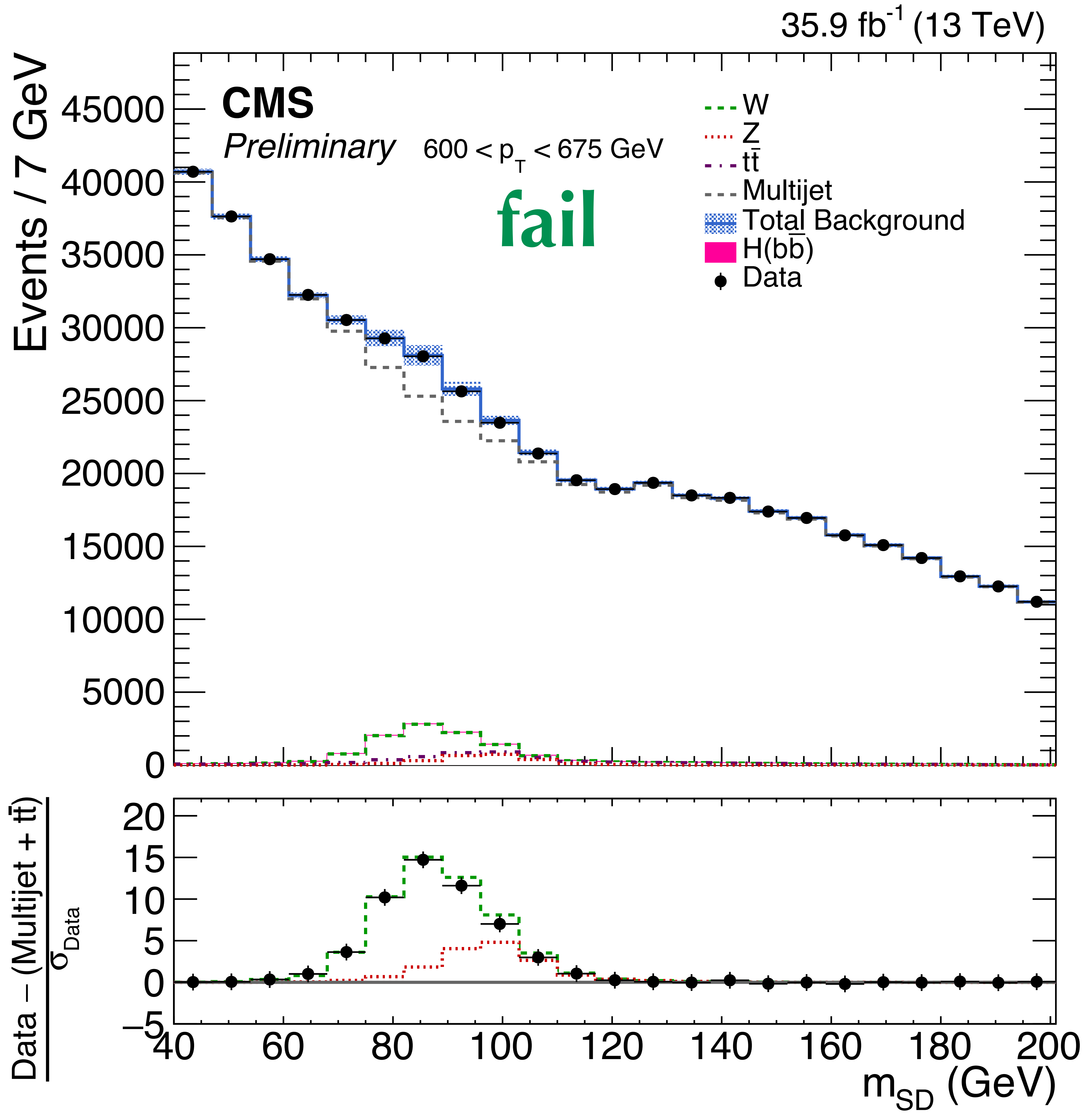
Signal+background fit

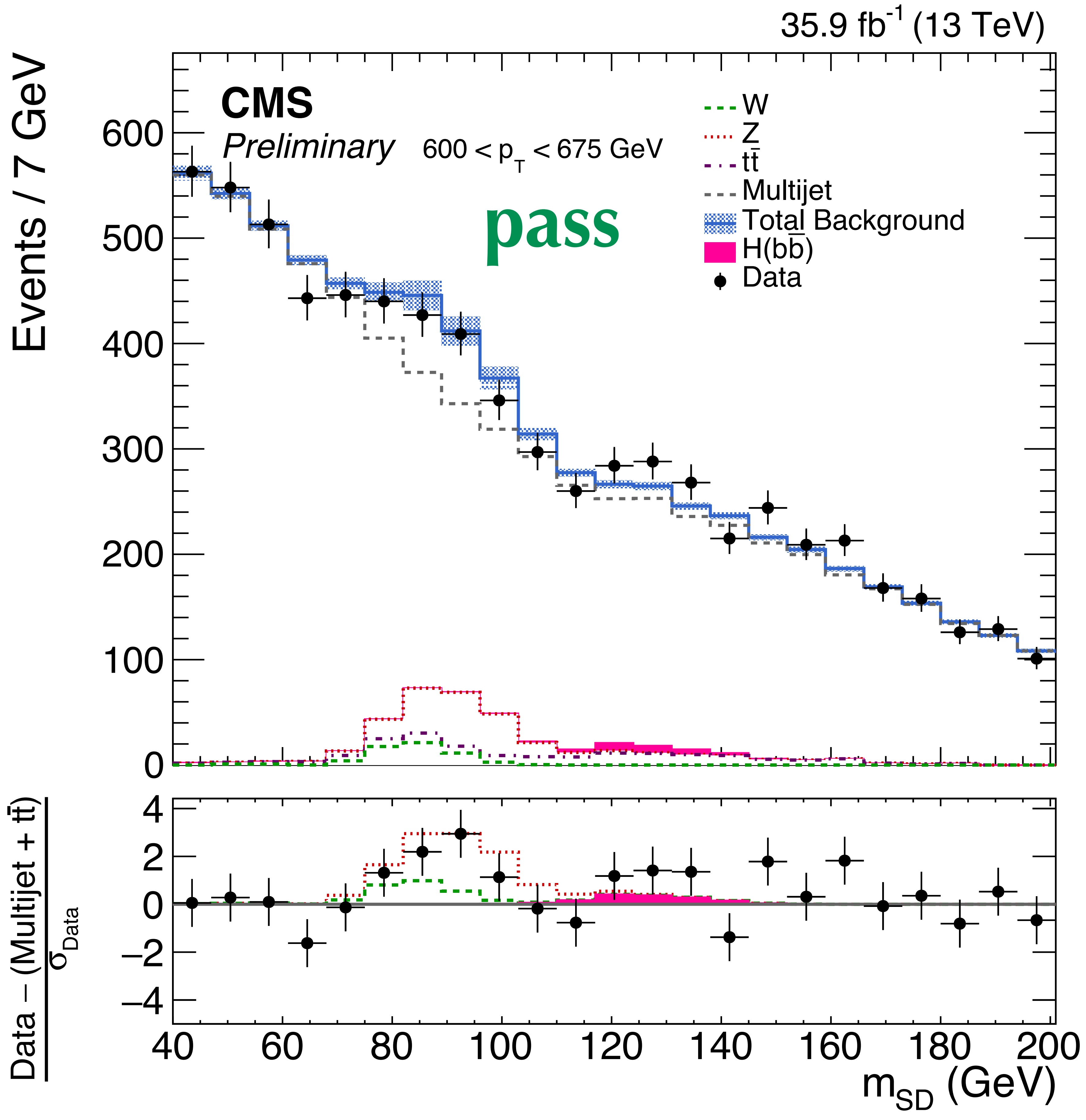


Signal+background fit

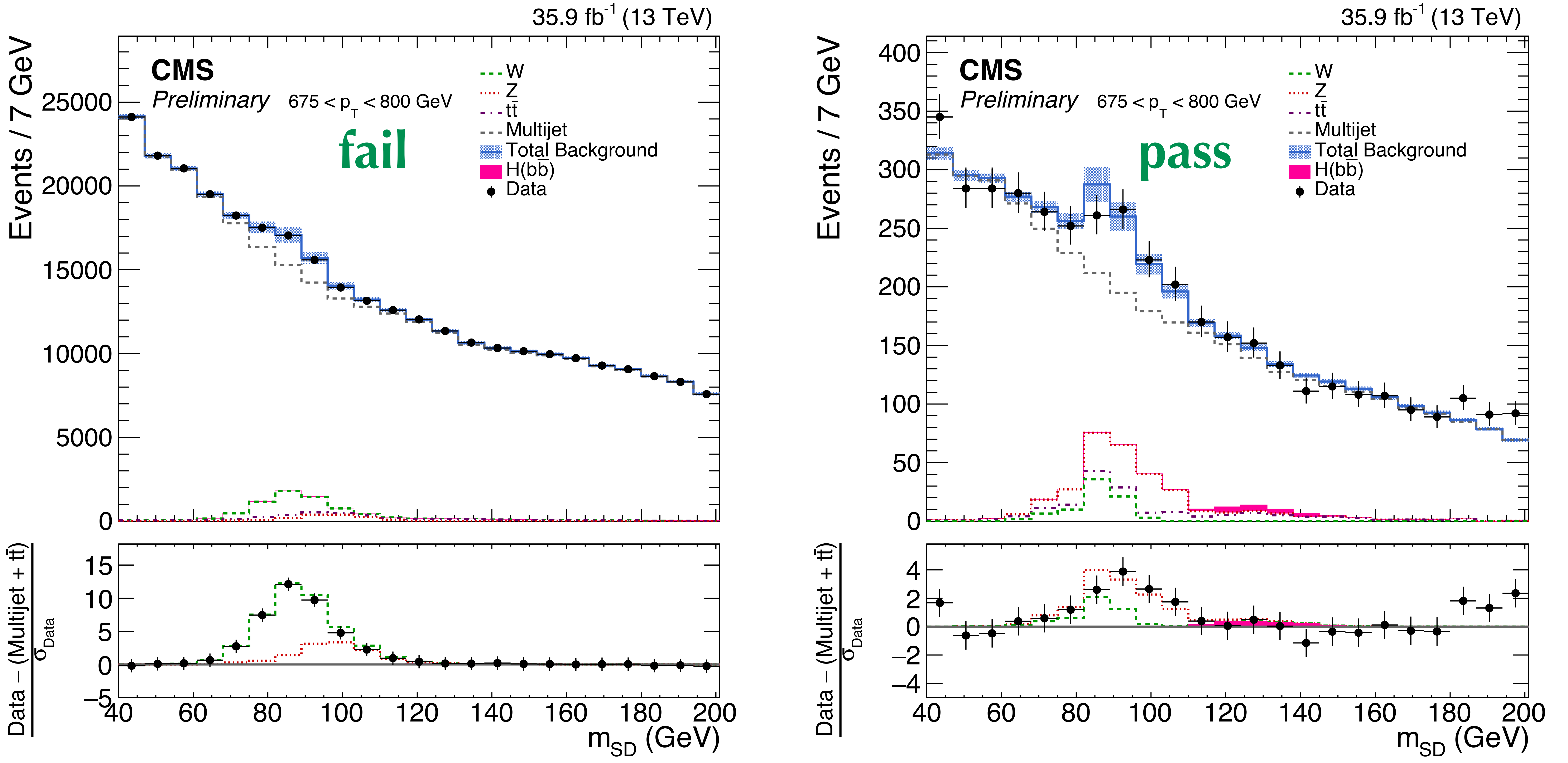


Signal+background fit

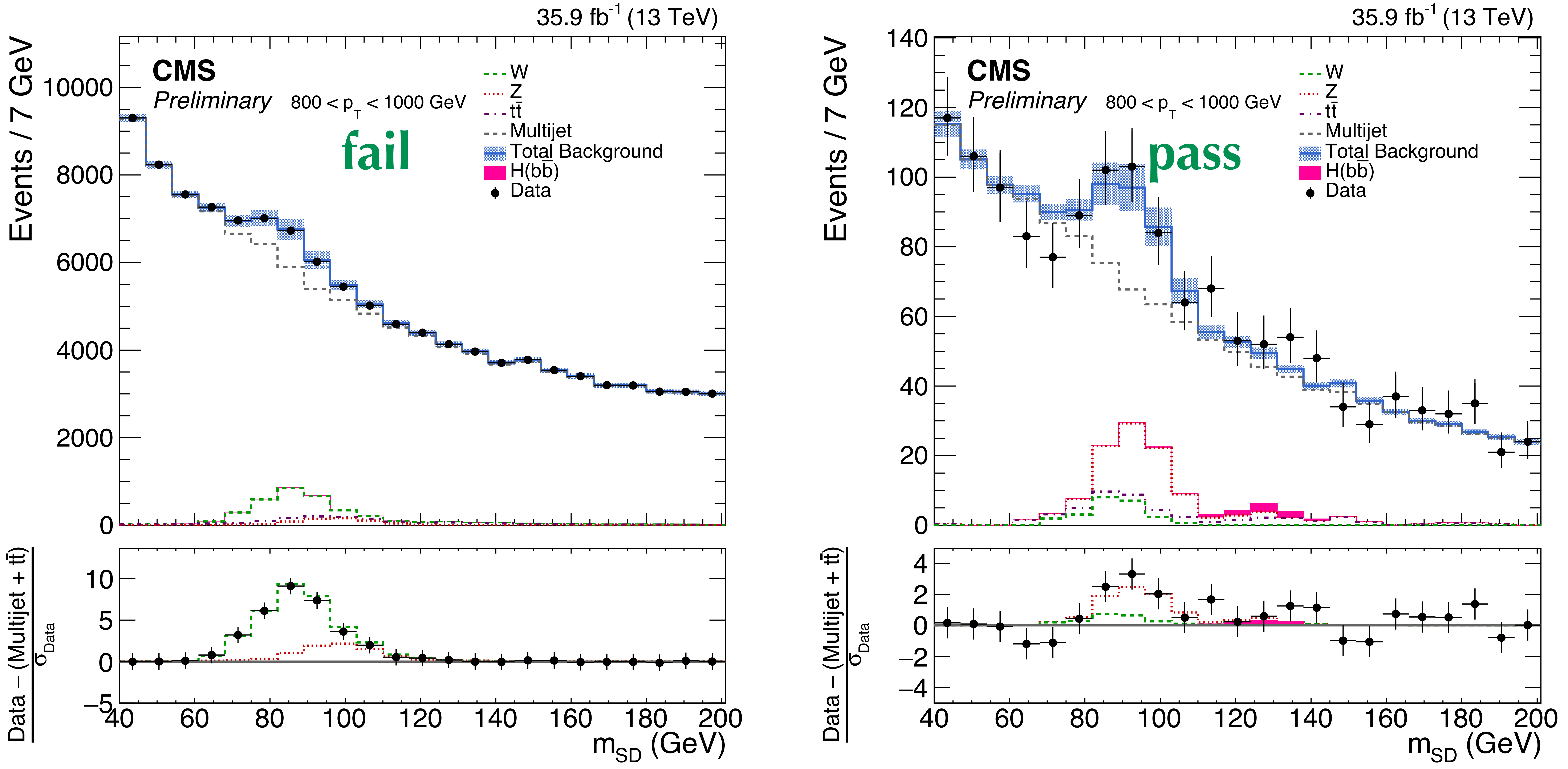




Signal+background fit



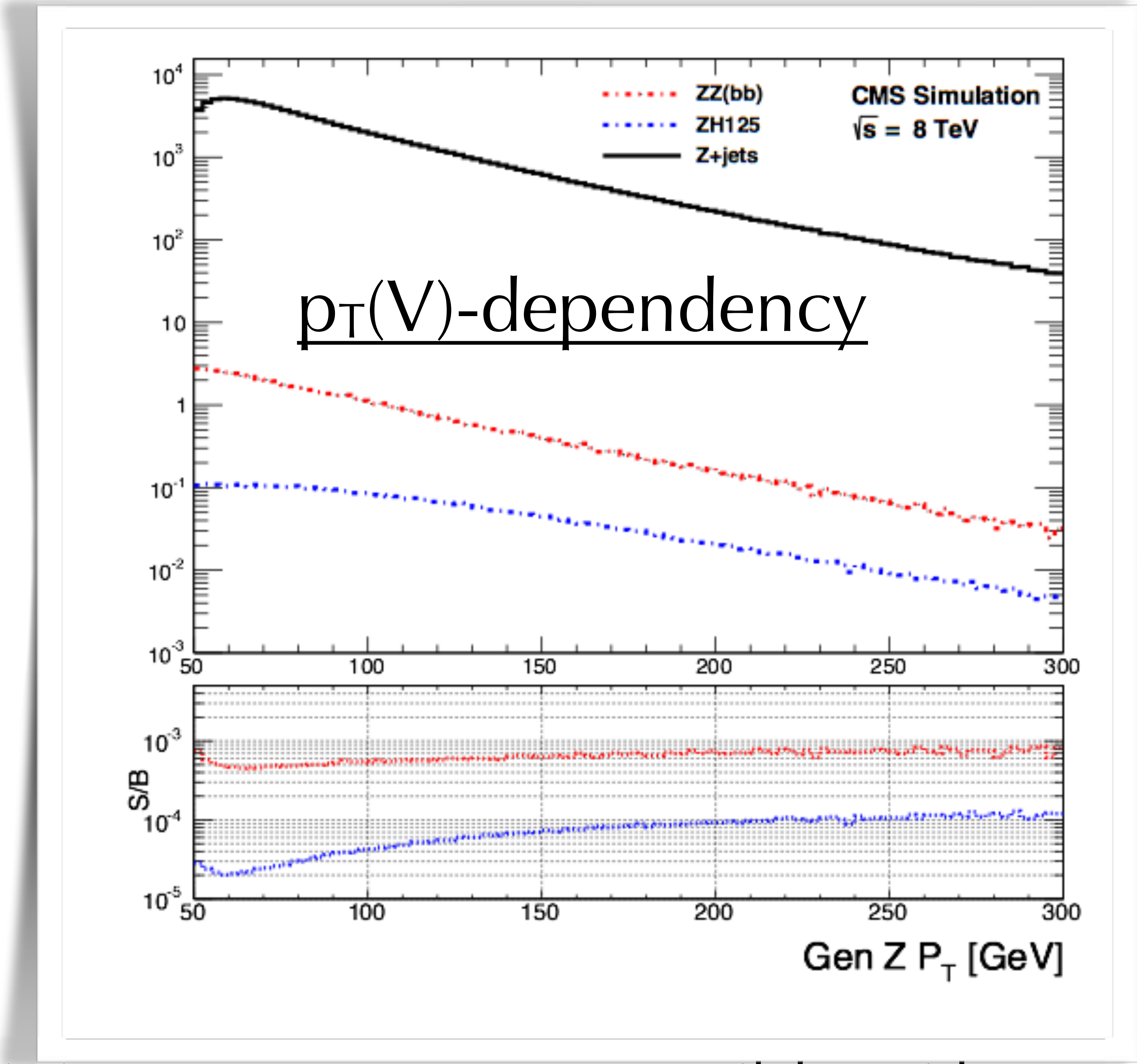
Signal+background fit

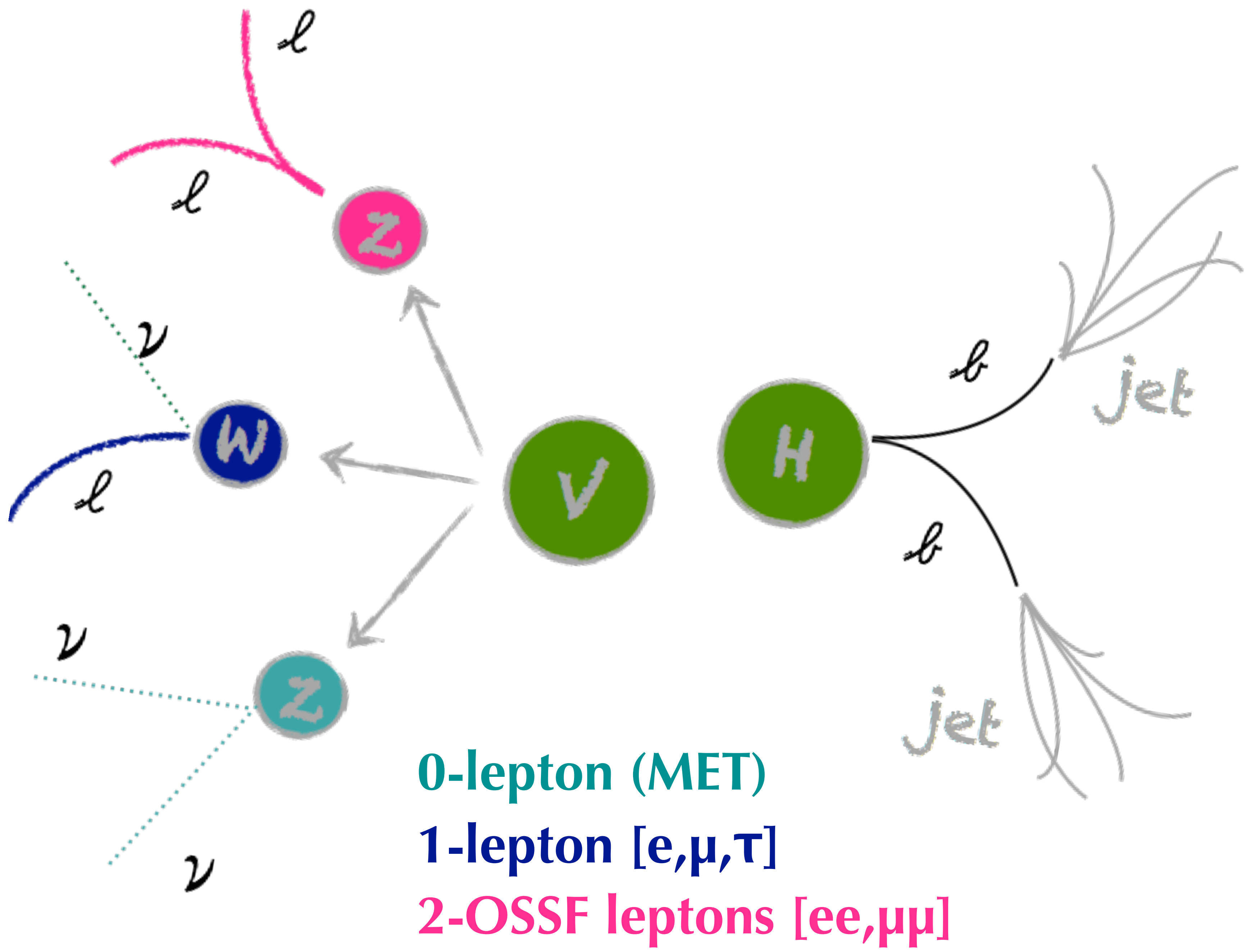


Phys. Rev. D 89 (2014)

VH, Event Topology

- $H \rightarrow b\bar{b}$ at LHC is searched in events where H is produced in association with a W or Z boson with high boost (~ 100 GeV)
  - events are triggered by the leptonic decay of the W/Z (e, μ, MET)
  - multi-jet QCD background is highly suppressed

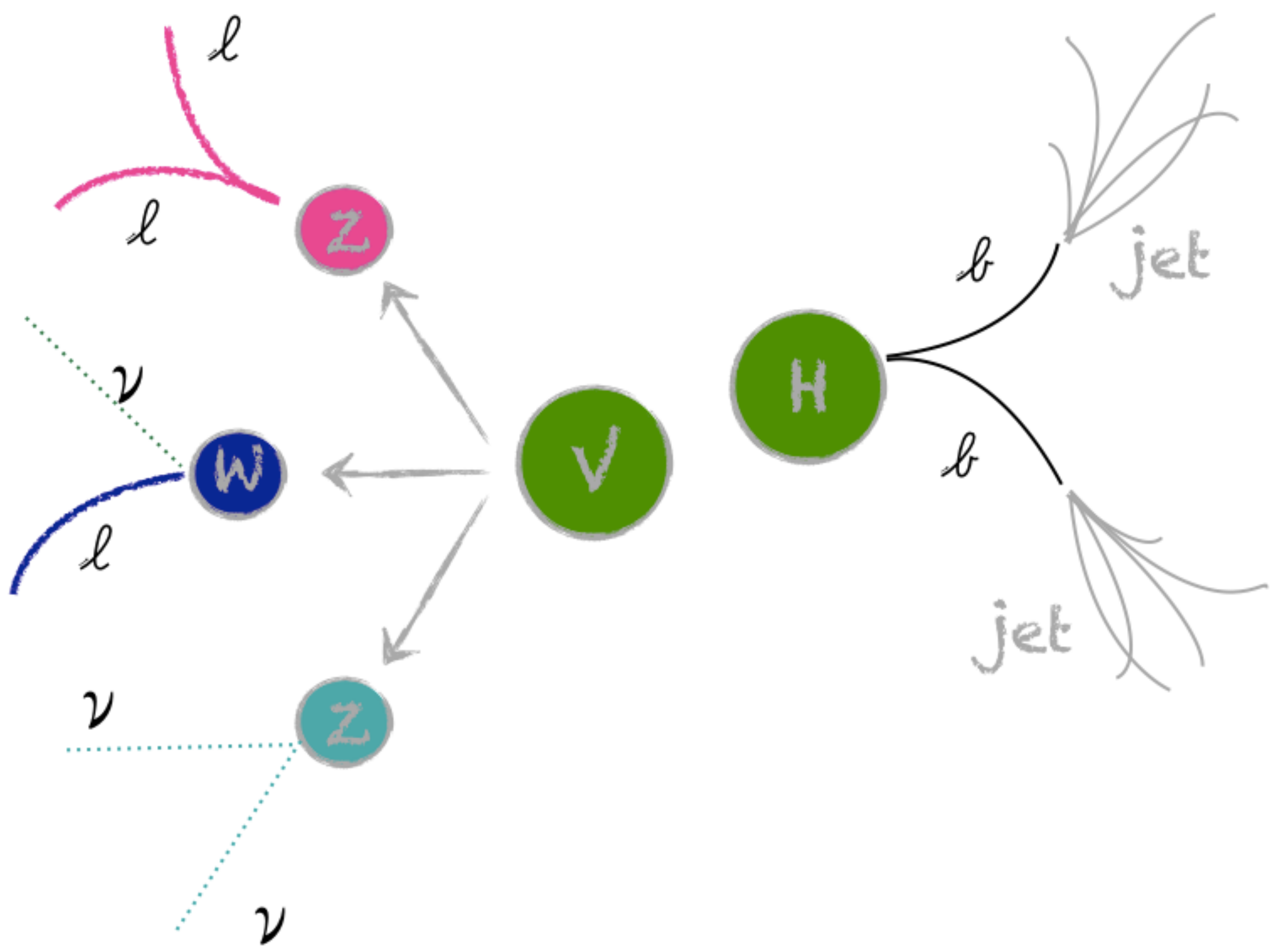




Quick look at the backgrounds

VH example

signal

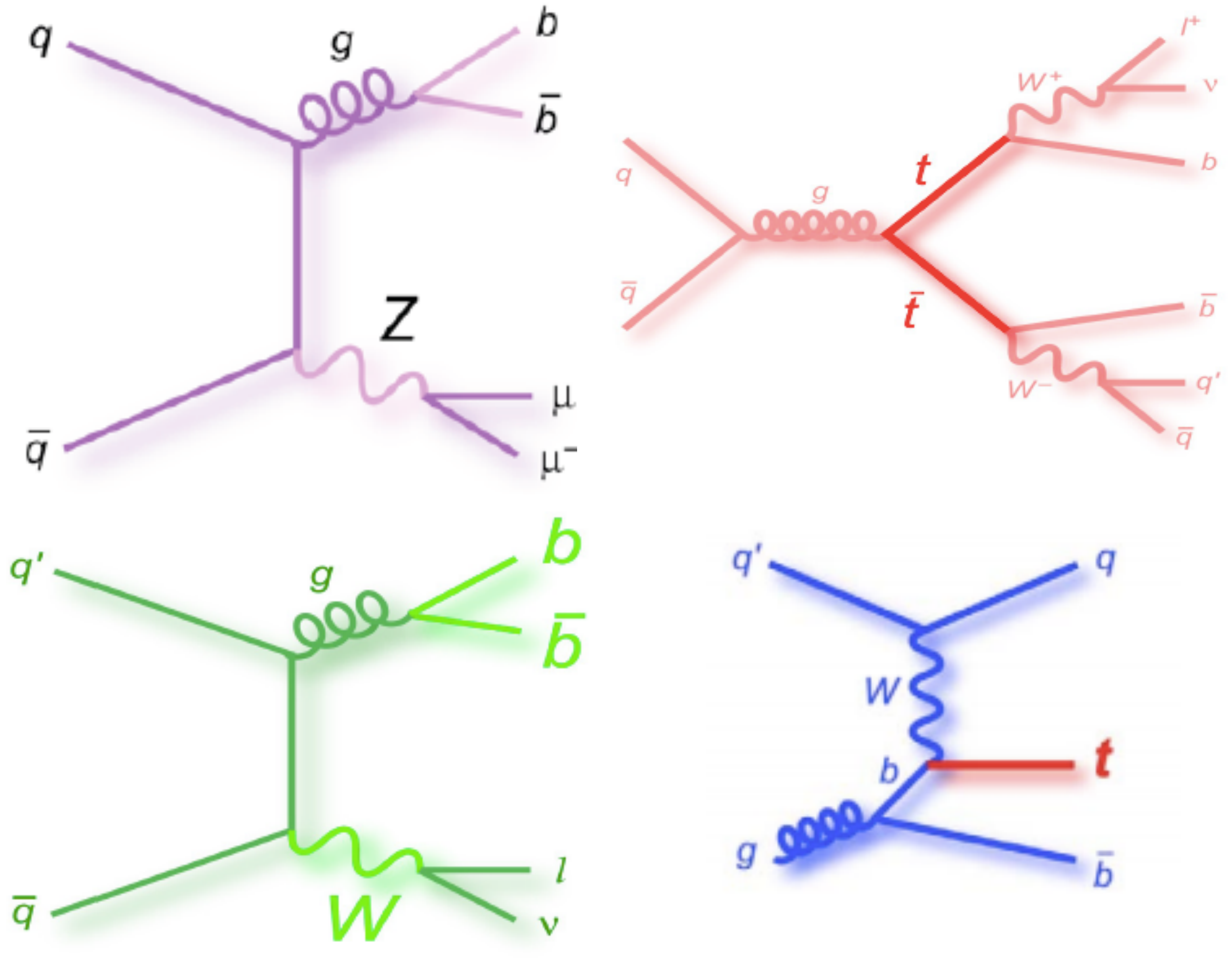


0-lepton (MET)

1-lepton [e,μ,τ]

2-OSSF leptons [ee,μμ]

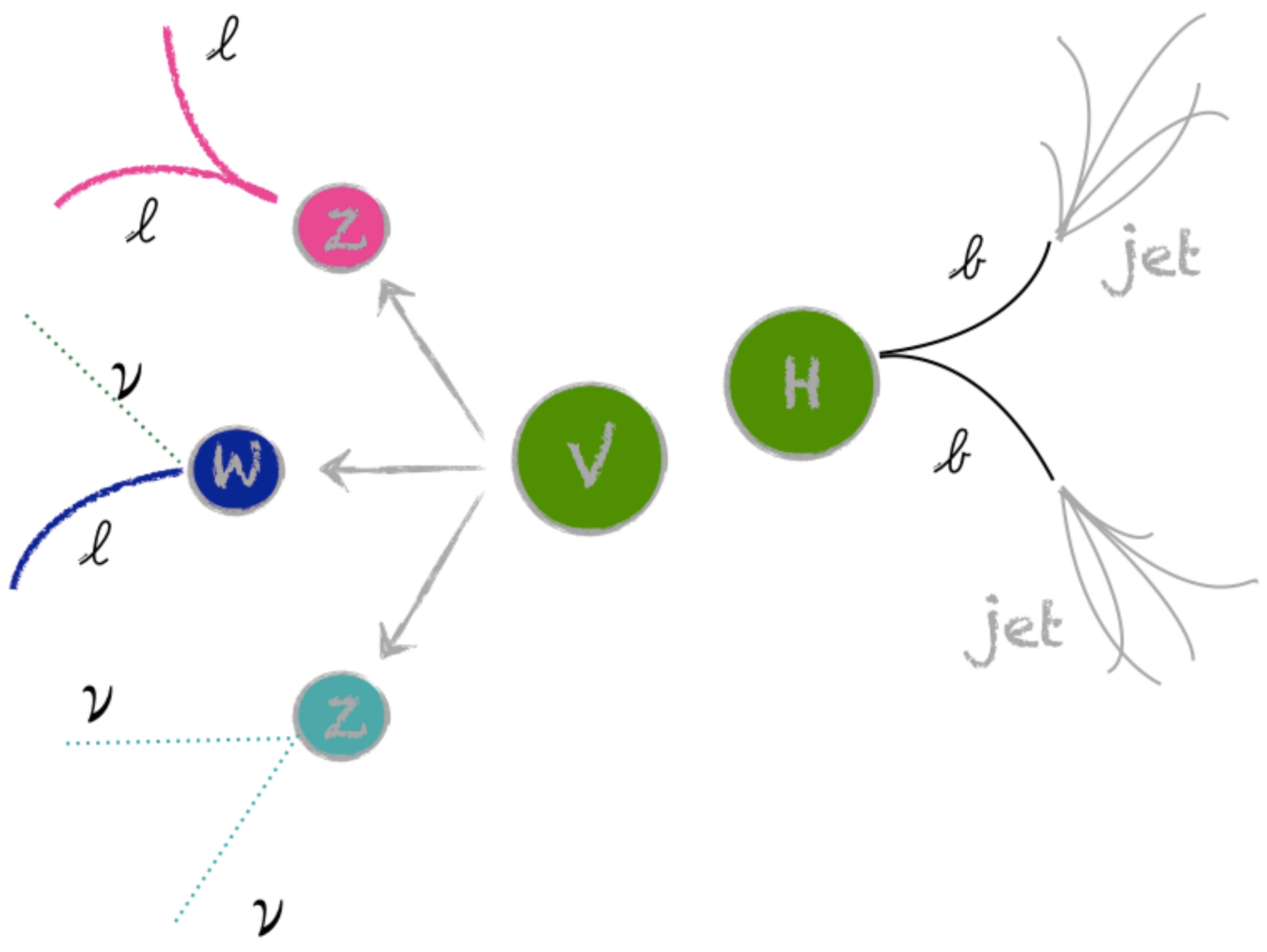
irreducible backgrounds



Quick look at the backgrounds

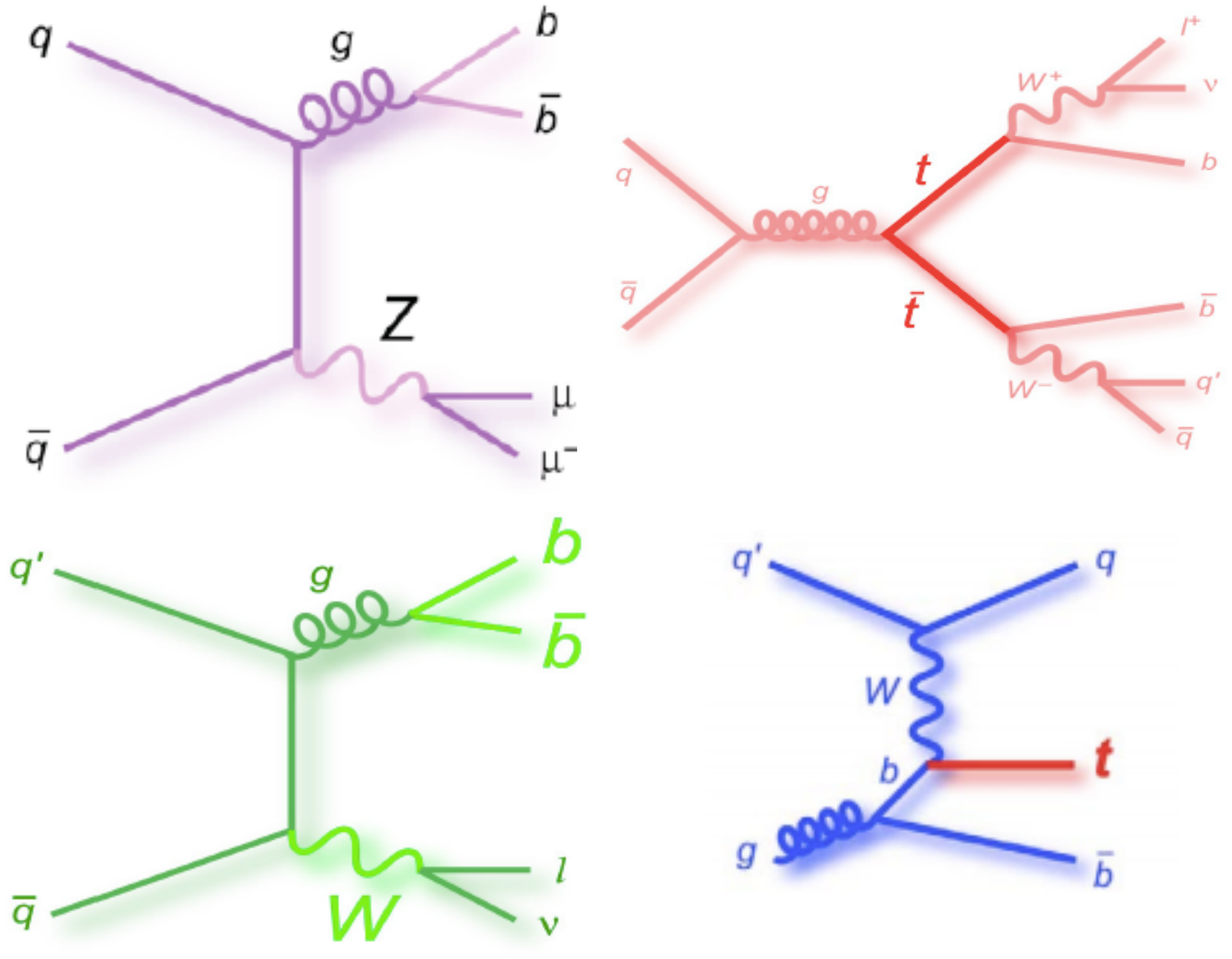
VH example

signal



0-lepton (MET)
1-lepton [e,μ,τ]
2-OSSF leptons [ee,μμ]

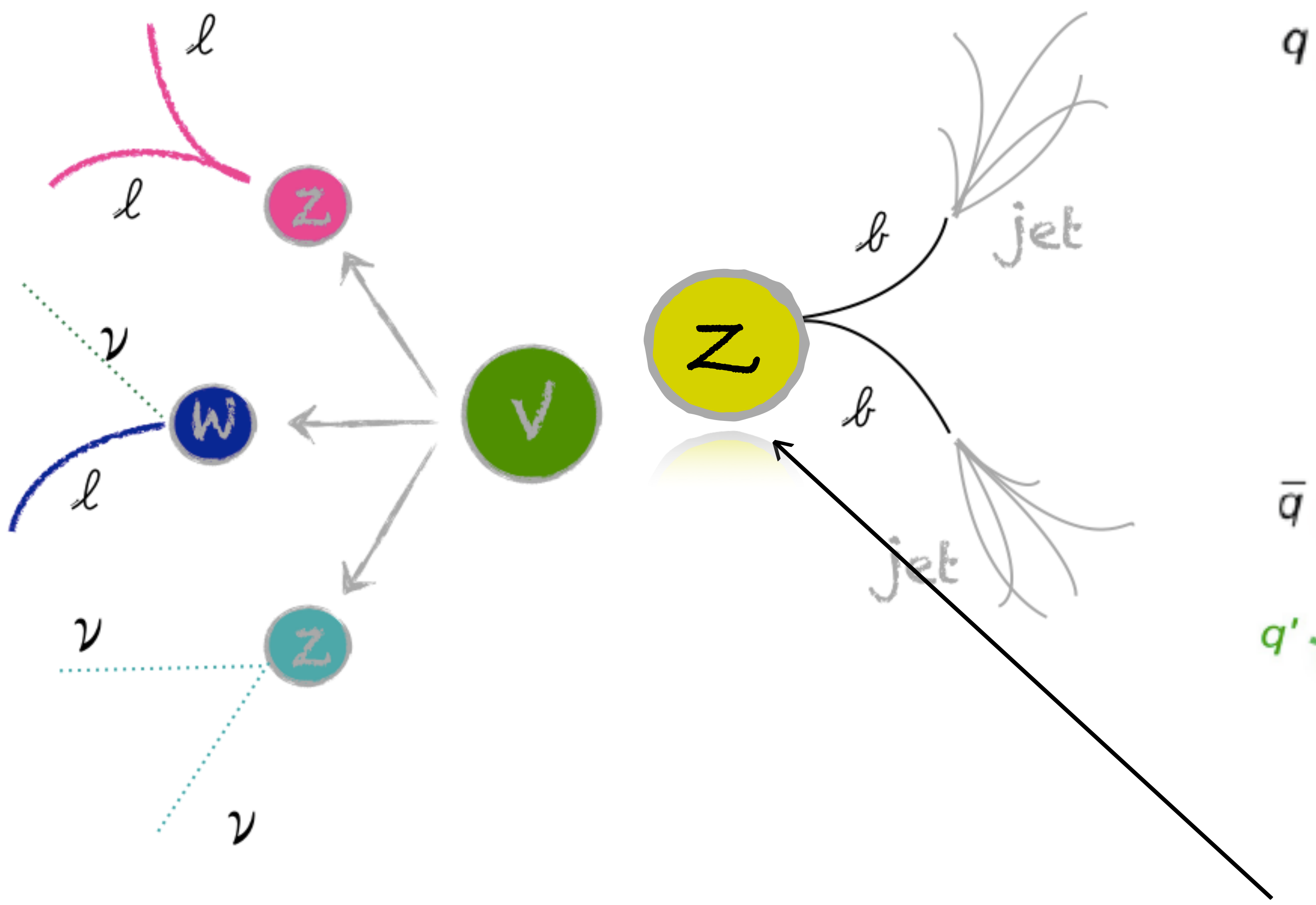
irreducible backgrounds



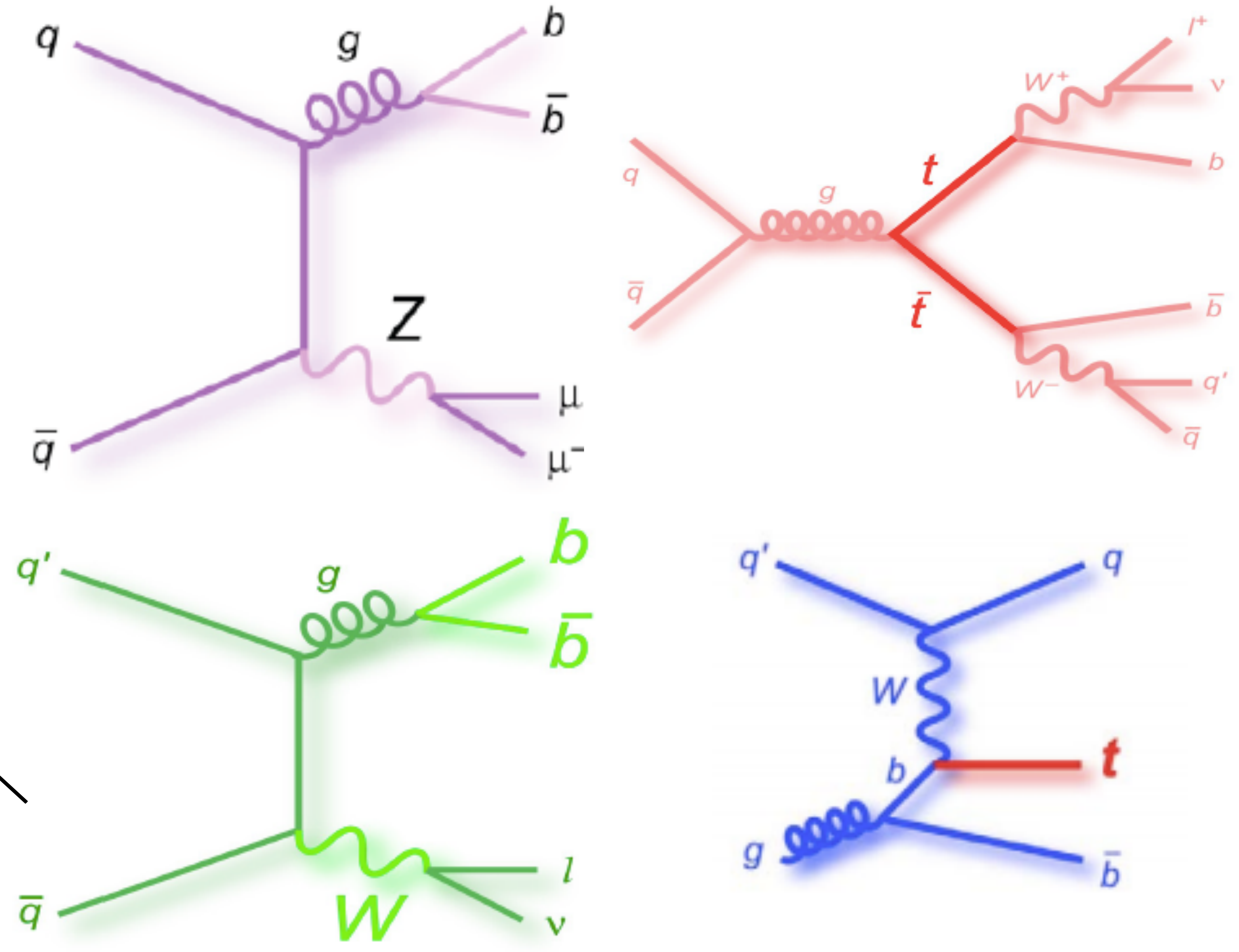
and diboson, of course

Quick look at the backgrounds

VH example



irreducible backgrounds

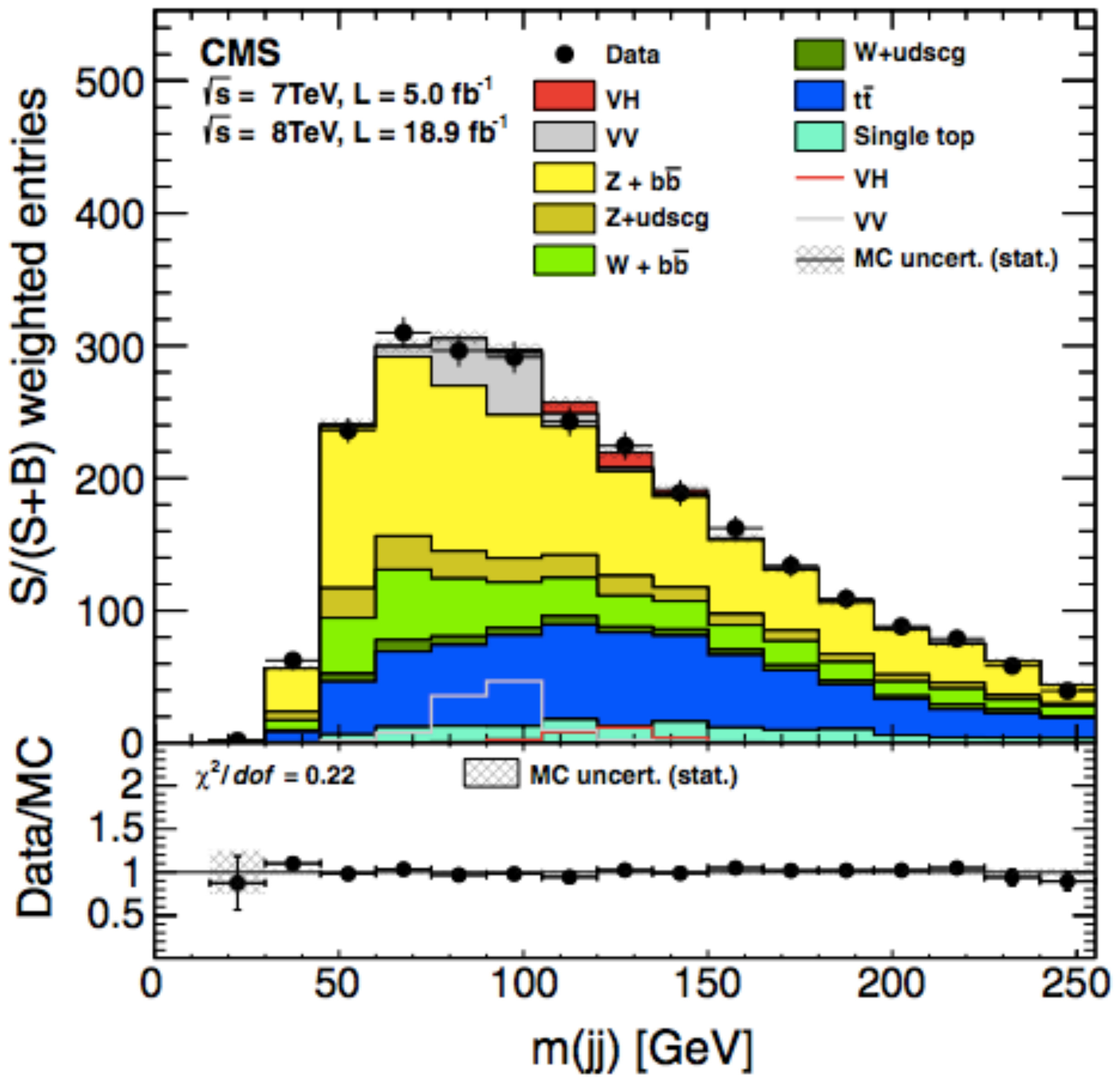


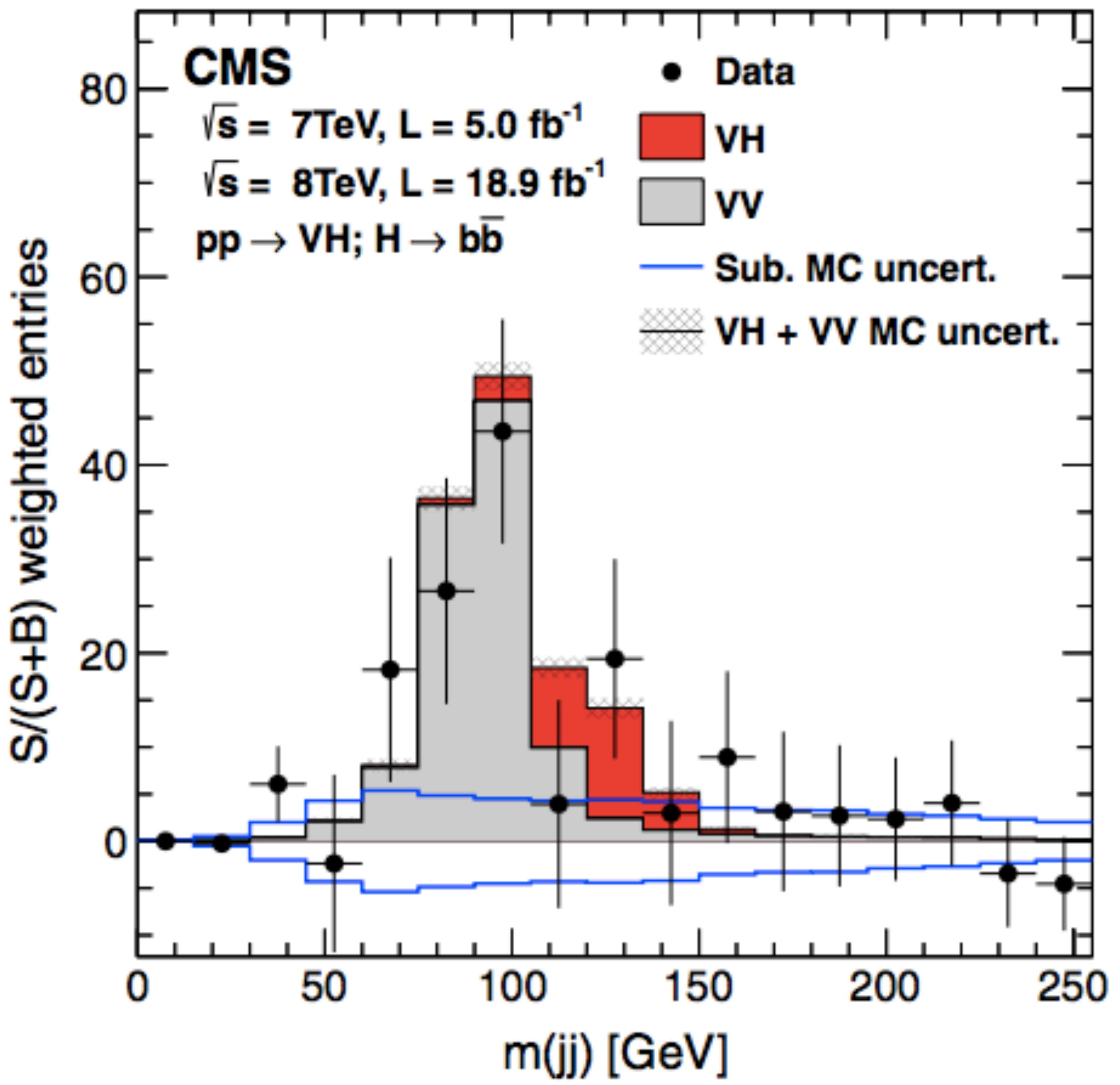
0-lepton (MET)
1-lepton [e,μ,τ]
2-OSSF leptons [ee,μμ]

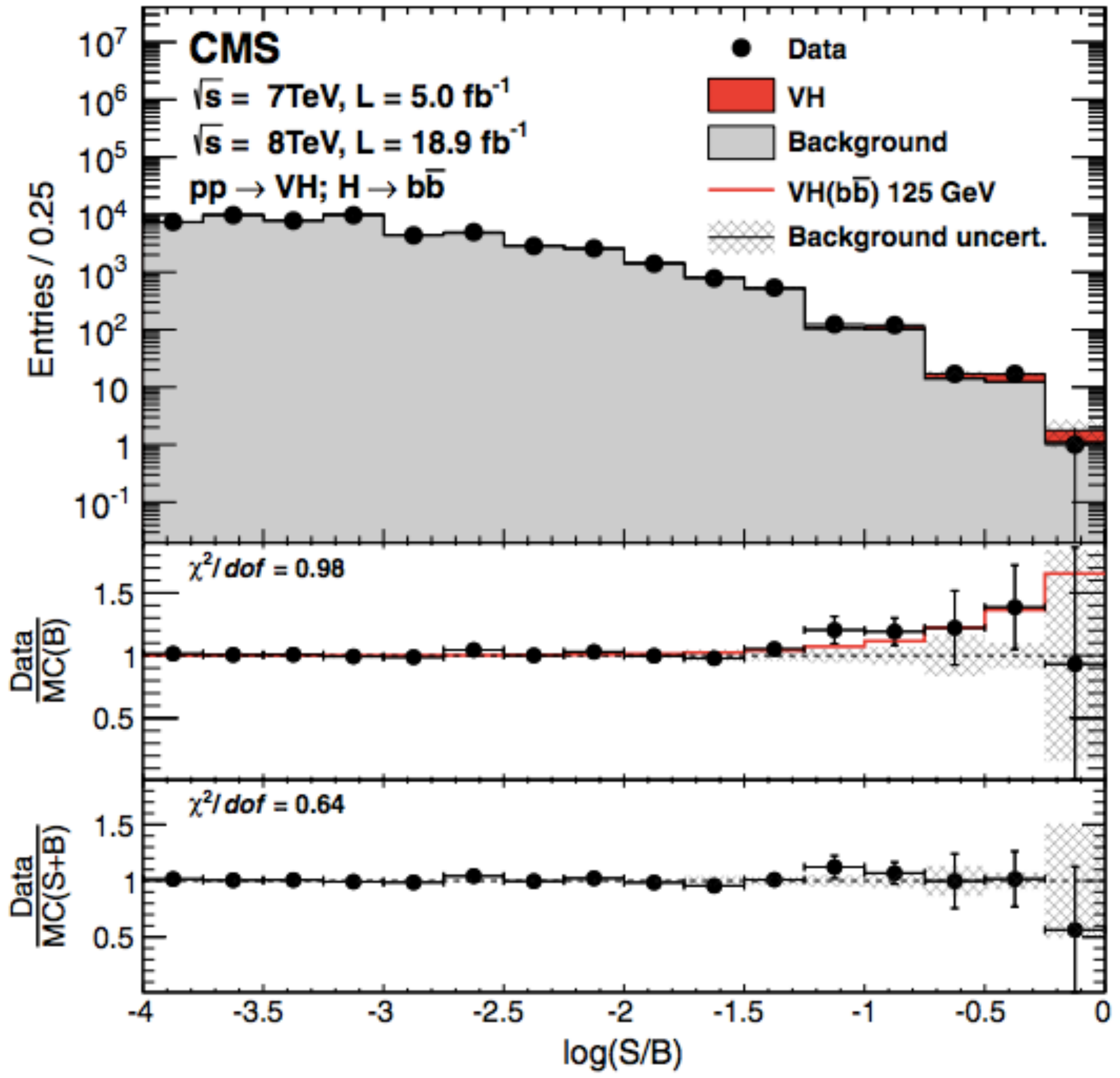
and diboson, of course

Phys. Rev. D 89 (2014)

VH, Analysis Strategy







Key points:

1. Extract normalization for the dominant backgrounds from the data V+0b/1b/2b and top pair production
2. b-jet energy specific corrections (**regression**)
3. A multivariate analysis, BDT

Phys. Rev. D 89 (2014)

VH, Results

VH($b\bar{b}$) reported an excess of 2.1 σ in agreement with SM H expectation at 125 GeV

- ✓ $\mu = \sigma/\sigma_{SM} = 1.0 \pm 0.5$
- ✓ All modes are compatible
- ✓ most sensitive result, to be compared to

CDF $\mu = \sigma/\sigma_{SM} = 2.5 \pm 1.0$ **D0** $\mu = \sigma/\sigma_{SM} = 1.2 \pm 1.1$

D0+CDF $\mu = \sigma/\sigma_{SM} = 1.95 \pm 0.75$